

Higher education in trouble

The debate about the British government's plans for its universities will soon begin in earnest, but the objectives of the intended changes are far from clear.

WITH the British parliament back from its holidays this week, the government's plans for education are back on the agenda. Or some of them are. The House of Commons has already begun going through the huge education bill clause by clause. Soon it will have sunk its teeth into the most contentious parts of Mr Kenneth Baker's bill — those giving schools the right to opt out of local authority control and parents the right to choose where they send their children. What with a budget looming up, there will not be much serious discussion of the bill's proposals on higher education until it gets to the House of Lords early in the summer. Only if it is then amended in ways the government considers unacceptable is it certain that the half of the British parliament enjoying electoral legitimacy will explicitly consider the most radical changes this century in British higher education.

That, by itself, is a striking measure of how times have changed in Britain. The universities in particular have been politically friendless for a long time now — at least for a decade, perhaps since the formal designation of the polytechnic system ten years earlier. Those who doubt that should dig out the prospectuses offered by the opposition parties at last year's general election. None promised the restoration of financial or moral independence to the universities, but each offered a different recipe for bending higher education to different national objectives. The opinion that, out of self-preoccupation and indifference, British universities have failed the national economy is widely held. The fact that the complaint is untrue is almost neither here nor there, for it is the motivation of the parts of the Education Reform Bill touching higher education.

A curious situation has arisen that will, it is hoped, be explored at the seminar *Nature* is holding next week in London (see box). Mr Kenneth Baker, Mrs Thatcher's third Secretary of State for Education and Science (before Sir Keith Joseph was a Mr Mark Carlisle, remember?) is potentially the most interesting yet, but has been in the job for only a year. Not surprisingly, his education bill brings together several strands of government opinion — suspicion of elected local authorities (opting out and the abandonment of the pretence that polytechnics are run by local authorities), populism (parental choice), the maintenance of standards (periodic achievement tests in elementary and secondary schools), public accountability for the spending of public money (the tight accounting rules by which universities will live under the bill, if enacted) and the belief that business, the engine of economic growth, knows best (whence the stipulation that the Universities Funding Council will have a businessman, not yet selected, as its chairman). Yet despite the explicitness of the new recipe for higher education on procedural matters, neither the bill nor the white paper which preceded it

last summer adequately describes the objectives of the purported reforms. Even the British government's critics are forced to guess at what may be in its mind.

The most charitable opinion is that the British government honestly believes it can further help to make Britain a more prosperous place (its achievements in that regard by other means should not be sniffed at) by making higher education walk a more clearly defined line. But with what objective? The British government does not acknowledge the peculiar character of higher education, which is the process of educating adults destined demographically to outlast their teachers and, on the side, of sustaining scholarship (by research). The government agrees with Britain's academics that the two activities are essential (but not necessarily that they are also, at some level, inseparable). The weakness of its philosophy of reform is that it seems to regard each of these functions as a commodity whose supply must be increased. Britain needs more educated people, so they must be trained, while research has economic value; if it happens that it cannot be exploited domestically, at least the patent rights may be saleable.

But higher education is not just another kind of business. The supply of its products, in the British government's idiom, may be only crudely related to the demand (or need). Thus Britain is in desperate need of more and more skilled technical people, but will the young devote themselves to a profession sullied in public reputation by the shrunken payrolls of the past few years? And while researchers working in partnership with industry may stand a chance of making a fortune now and again, what is to happen to those with an intellectual interest in other things? The other implicit misconception underlying the bill is that the universities, dependent as they are on public funds, should be compliant with the undefined inclinations of their sponsors, is equally mistaken. It is a plain truth that the most socially valuable of universities are often the most heterodox. (Where would the Thatcher government be if it had not been for the monetarist heterodoxy of the University of Chicago in the 1960s?) Higher education will not be worth having if it ceases to be subversive of present ideals, which is the real reason why its institutions need to be structurally autonomous. The British government will not help itself by the centralism around which its reforms are built.

On the bill, academics might take this line. The pinpricks, such as the proposal to extinguish academic tenure, are not as important threats to academics' freedom as they have been made out to be; academics could be fired only when the need for their services had disappeared, and universities (not the government) would be the judges of when that circumstance had materialized. The central issues are the provisions that the funding council may attach conditions to the grants it makes to particular universities and that the government may direct the funding council to attach conditions. Those who know the way the present system works are in two camps: some say that even present customs give the central government absolute power which it dares not exercise for fear of the fuss that there would be. Others, however, note that the mere existence of a constitutional right of interference from the centre gives civil servants a sense of mission to interfere. The serious threat is that British

The education bill: reform or otherwise?

A SEMINAR on the implications of the Education Reform Bill organized by *Nature* will be held in London on the afternoon of 19 January. The principal speaker will be Sir Mark Richmond, Vice-Chancellor of the University of Manchester and Chairman of the Committee of Vice-Chancellors and Principals. Those wishing to attend should apply for tickets and further information to Mary Sheehan (01-836 6633 extension 2223). □



higher education will be run by the British civil service.

Mr Baker keeps saying that he has no intention of using the powers his bill will give him, in which case he will presumably volunteer amendments to the present text that will still the anxieties on this score that have been voiced so far. Academics might usefully bend their energies to devising suitable forms of words. The other structural weakness of the new arrangements is that they rely on the recruitment of a small army of successful business-people farsighted enough, and with sufficient spare time, to labour at the administration of the new system. It is for the British parliament to find out whether the supply of these people will meet the demand Mr Baker plans to create.

The pity is that the proposed reforms, as they are described, will not advance the most respectable of the British government's intentions — that of making the British labour force better suited to what the world will be like next century. Indeed, by further distracting academics (already too committee-bound) from the jobs for which they are employed, the process of reform will be in part self-defeating. The educational matters left untouched by the bill as it stands concern the ways in which students are to be prepared at school for higher education and the manner (and the length of time) of their careers in the tertiary sector. Although the participation of young people in Britain has risen (to 16 per cent), that is only half of the corresponding ratio in Japan. And while the government of France (see page 104) is just now wrestling with some of the problems on Mr Baker's agenda, no harm would come if the British parliament broke, on this occasion, with its habitual practice of supposing its own problems to be unique. □

Less to spend

The need to cut the US budget deficit will have significant consequences for research.

THE pre-Christmas panic to reduce this year's federal budget deficit by at least the \$23,000 million required to avoid automatic cutting will bear significantly on the spending plans of the civilian agencies in Washington which support research. So much (see page 101) is now clear. Perhaps the greatest disappointment is that the National Science Foundation (NSF), offered at the beginning of 1987 the prospect of growing to twice its then size over five years, will fall significantly below that growth curve. It will be a matter of great concern to see whether the goal is shaded in the budget for the next fiscal year (beginning in October), now only a few days from publication. It will do the administration great credit if it does its best, at the beginning of the last of the financial years it can hope to markedly to influence, if it begs and borrows from elsewhere in the federal budget so as to give NSF a chance to resume its growth. There will be more honour in the history books (and more profit in the balance of external trade a few years from now) for a \$10 million spent on civil research than \$100 million on new weapons systems.

But nobody should pretend this will be easy. Next year's budget will be even tighter, while it is far too soon to tell whether the world's financial markets will allow the administration to get away with the modest further reduction planned for 1988–89. Even at the time of the month-long congressional haggle about deficit-reductions after the stock-market tumble of 19 October, it seemed plain that barely staying within the legal limits would not suffice. Events since then have confirmed that suspicion. On balance, there is a high chance that stability will not come about without a more marked reduction of spending. That is on the assumption that the administration will not voluntarily choose the best course of increasing taxes modestly. But that is a safe bet in an election year, but it is entirely possible that such a step may be forced upon it, even with so little time left to go before November. □

Are germ-lines special?

The conclusion that genetic manipulation of the germ-line must be outlawed may be too hasty.

THERE is a great risk that we shall all drift into a muddle over the prospect that genetic manipulation of the nuclei of human cells might be used as means of counteracting the inherited effects of genes that happen to be deleterious. In the United States, Mr Jeremy Rifkin continues his campaign for the sanctity of all species' germ-lines. The British government's white paper on the general subject of what is called the new embryology, likely soon to be followed by draft legislation, is just as explicit as Mr Rifkin; acknowledging that the prospects of germ-line genetic manipulation are remote, the white paper nevertheless says "It is a procedure which society would clearly regard as ethically unacceptable, and the Bill will prohibit it". Readers of *Nature* will have noted that Professor David Weatherall made much the same point (331,13; 1988), in a comment on a paper from Dr Richard C. Mulligan and colleagues at the Whitehead Institute, when he said that germ-cell manipulation should not be applied to human beings "because it would alter the genetic make-up of subsequent generations". That, of course is true, and the opinion is widely held. But is interdiction justified on that ground alone?

In spite of all the valuable, if anticipatory, attention lavished on the study of genetic manipulation over the past few years, several important and ultimately practical questions have been ducked. One of these is precisely the distinction between the still remote treatment of inherited genetic diseases by manipulating the genes of somatic cells and the still more distant prospect of manipulating the genomes of early embryos so as to avoid the possibility that unwanted genetic diseases will emerge. Why does the British government suppose that to be a "procedure that society would regard as ethically unacceptable"? If, indeed, it were the case that such a technique might utterly remove from a family the risk of carrying genetic haemophilia (and there is at present not the vaguest notion of how that might be done), might there not be circumstances in which society would welcome the development? Royal families with haemophilia would no doubt have jumped at the technique.

The reason why it is commonly supposed otherwise is not, of course, that breaking the transmission within families of known deleterious genes would be unacceptable, but that it is easy to think of ways in which a technique for manipulating the genetics of the human germ line could have horrendous consequences, by design or from mere ignorance. Engineering families of people with looks of the kind qualifying them to be film stars would no doubt appeal to some, but would probably be unproductive and certainly a waste of time. The knowledge that the genetic component of IQ must certainly involve many genes would similarly not deter clever people from wishing to perpetuate their own kind. There are nightmares to be conjured out of the contrary possibility, that families of unintelligent or dependent people might be founded deliberately, although it is difficult to see what the benefits of that might be. Indeed, as a little reflection will show, it is very difficult to see how genetic manipulation of the human germ-line, supposing it is ever practical, could profitably be put to malign application.

Is it sensible, in the circumstances, that potentially beneficial uses should be set aside without discussion? In the still-brief history of the public discussion of the uses that may be made of new techniques in cell biology, great benefits have followed from people's willingness to consider the worst cases, and then to decide deliberately how they must be avoided. This is the spirit in which genetic manipulation has been, from the outset, approached. To follow a different course in this connection will have the effect of ensuring that those who may responsibly worry about what may emerge remain in ignorance, leaving the field to those who are less responsible. □

US agencies sift for wheat in 1988 budget chaff

- AIDS research boosted
- Some last-minute winners

Washington

THREE months after the fiscal year began, the US federal government has received its spending authority from Congress for 1988. But even now, weeks after President Reagan signed the continuing resolution that authorizes just about every expenditure the government will make until September when this fiscal year ends, the various federal agencies are combing through the thousand-page law, trying to determine what restrictions have been placed on the money they have to spend.

In a year of belt-tightening, science budgets did not make out too badly. The National Science Foundation (NSF) was knocked off its course towards a doubling of its budget in the next five years, but it nonetheless received \$1,717 million, 5.8 per cent more than last year, but \$176 million less than had been requested.

Funds for research were set at \$1,453 million, 3.3 per cent more than 1987. Exactly how many of the new science and technology centres can be funded with the money available and how many individual grants can be awarded will be decided by the end of this month.

Although Congress has left to NSF most decisions about where to put its money, the agency was required to give \$1.5 million to a supercomputer centre that offers unique resources not available from those supercomputer centres already supported by NSF. This money will probably go to the University of Minnesota, which has the only Cray 2 supercomputer at an academic facility. Also mandated by Congress is \$1 million to IIASA, the International Institute for Applied Systems Analysis located in Laxenberg, Austria. In the past, the National Security Council has blocked grants to IIASA because of concern that sensitive information was leaking to the Soviet Union through IIASA.

The National Institutes of Health (NIH) also received their customary favourable treatment from Congress. The total sum amounted to \$6,667 million, up approximately \$400 million from last year. That amount will allow NIH to make between 6,000 and 6,100 new and competing grants. The budget for the Centers for Disease Control in Atlanta will be \$772 million, 75 per cent more than requested by the Reagan administration.

Congress also left most decisions on spending up to NIH, but did insist that \$22 million be used for facilities devoted

to coping with the public health crisis brought on by AIDS (acquired immune deficiency syndrome). Total AIDS funding to NIH was \$448 million, nearly half the total of \$906 million given to the Department of Health and Human Services for AIDS activities.

Congress has begun to show interest in the project to map and sequence the human genome. Federal money will come chiefly from two sources; approximately \$17.2 million for mapping efforts from NIH, and \$12.5 million from the Department of Energy (DoE) for its activities, including work already started at three national laboratories (see p. 103). NIH will also continue to support individuals working on efforts related to the project.

DoE has other big projects to contend with. Its 1988 budget includes some \$25 million for the Superconducting Super Collider. The agency expects to pick a site for the proton-proton collider this summer. But DoE has some tricky juggling to do. Although the Office of Energy Research received some \$2,050 million to spend, nearly 5 per cent of that total will have to be given to 'earmarked' funds, money that Congress specifically directs for spending in particular ways. This money often goes directly to universities, bypassing any form of merit review. The DoE basic energy science has been a favourite place for these earmarked funds to be hidden in the budget, as has the Agricultural Research Service's budget (see box). Although some universities have tried to resist the temptation to make such direct appeals for facilities through political muscle, Robert Rosenzweig, president of the American Association of Universities, expects the practice will increase unless the government seriously tries to establish a pool of money for new facilities for which universities can compete.

The National Aeronautics and Space Administration did fairly well in the 1988 budget. The space station will receive \$425 million for the year, plus some \$80 million carried over from 1987 money. But that is still \$262 million less than the agency had been seeking. Congress made smaller cuts in budgets for planetary science missions, Earth science missions, and its sounding rocket and balloon programme, providing the requested amount for the Hubble Space Telescope only after a fight. Work will also be able to proceed on a fourth shuttle orbiter to replace the Challenger.

Joseph Palca

More money in the pork-barrel for some

Washington

IN a few instances in its 1988 budget deliberations, Congress saw fit to dole out more money than requested by the Reagan administration. Research on AIDS (acquired immune deficiency syndrome) and scientific and engineering education are two areas that Congress felt deserved special financial attention. But sometimes money shows up in budgets at the last moment, and its appearance is based more on the political skill and muscle of a particular state's elected officials than on national priorities.

The facilities account of the Agricultural Research Service (ARS) is an example. ARS, the research arm of the US Department of Agriculture (USDA), received \$540.7 million for its 1988 budget, some \$6.3 million more than requested. But for facilities, ARS requested only \$1 million for a national seed storage facility in Fort Collins, Colorado. But after the ARS budget emerged from the House of Representatives Appropriations Committee, that amount had jumped to \$14.7 million, with money for research facilities at nine universities and two USDA facilities around the country.

Not to be outdone, the Senate raised the facilities budget to \$50.0 million, adding five universities and four state facilities to the list of recipients. But that was not the end. When the conferees met to iron out differences in the two appropriations, the account rose yet again to \$57.8 million, with one new university coming into money. Several facilities got big boosts in the conference. Construction funds for the Agricultural Research Facility at Oregon State University jumped from \$5 million to \$10 million and the Center for Applied Aquaculture in Hawaii jumped from \$850,000 in the House version to \$6.4 million.

Congress has frequently chosen the ARS budget to make appropriations that go directly to universities rather than through a merit review process. It is a process that most deplore, but few can resist.

Joseph Palca

Erratum

In the advertisement of Dr Stefan Marinov's International Congress on Relativity and Gravitation, to be held in Munich on 22–24 April, the following errors unfortunately appeared:

page 2, line 8 should read "...gauge transformations as conventionally taught. The potentials are absolute potentials."

page 2, line 21: "machine MAMUL" should be "Bul-Cub machine without stator".

page 2, line 32, "...name Einstein..." should be "...name of Einstein..."

The journal titles "*Philosoph. Mag.*" and "*Europhys. Lett.*" should have been italicized.

Job losses for cancer scientists at Ludwig institutes

London

WITH five new branches to its credit, the somewhat impenetrable Ludwig Institute for Cancer Research is closing down four old branches. About 80 scientists in Britain, Canada, Australia and Switzerland are affected. Members of at least two of the doomed branches are angered by the decisions, which fly in the face of scientific reviews of their work and have not been adequately justified.

The most recent, and smallest, of the branches to be given notice of closure is in Cambridge, England. When the branch's director departed two years ago, the Zurich-based Ludwig Institute decided against closing the branch and gave six-year contracts to its staff. Perhaps for lack

of space, a new director has not been found. Plans are afoot, however, to solve the space problem and set up an expanded Cambridge branch of the Ludwig in 1990. Despite positive review of their programmes of research, the three senior staff of the existing branch were given one year's notice, extended on appeal to two years, and have not been offered employment in the proposed new branch.

About 20 scientists, under Dr Bernd Groner, are affected at the Bern branch, set up only four years ago to focus on the molecular and cellular biology of breast cancer, as well as its therapy. The staff were given standard five-year contracts but, says Groner, it was his clear understanding that the contracts would be renewed given a positive scientific review. Without that assurance, says Groner, it would not have been worth building up the branch. A very positive review was notched up late in 1985. A further review was scheduled for early this year but was forestalled by the December decision to close the branch by the end of this year, a period of notice that Groner considers much too short for his staff to find adequate positions and funding in Switzerland. The only official indication of a reason for closure was that space was inadequate, at least for expansion. A problem had indeed arisen with the plan to move the branch into a new building being planned at the hospital to which it is attached, but there was no foreseeable problem in continuing in the existing space.

One-year notices of closure were also issued late last year to the Toronto and Sydney branches. The latter, directed by Dr M.H.N. Tattersall, has concentrated on studies of metastasis and resistance to chemotherapy. In Toronto, experimental and epidemiological studies on the causes and prevention of colon and breast cancer have been pursued under the direction of Dr W. Robert Bruce. The Toronto branch was founded in 1981 and the contract was extended in 1986. A statement issued by the Ludwig says that the additional years were provided for the completion of the research programme which will have been achieved by the end of 1988.

News of the apparently cavalier closure of at least two of the four branches has created concern among the staff of recently opened branches in London, Montreal, Stockholm and Uppsala. But the concern is counterbalanced by praise for the quantity and quality of support available for the new ventures. Moreover, the director of one new branch professes lack of concern. "It is the Ludwig's right to make whatever decisions it wishes; unfortunately it is not its style to explain why", he says, adding

that scientists need to become used to this style of employment and that a year is sufficient time for good researchers to find alternative employment.

Many of those involved with Ludwig affairs were prepared to say that it was not necessarily right to deal with the scientific staff in such an abrupt and insensitive way. Some suggested that the decisions were linked to an evolution of policy reflected by the recent appointment of Dr Richard Reitemeier and a new chairman of the board of directors. The possibility that the institute has over-extended itself financially with the new branches tends to be discounted. The institute's budget, which amounts to more than \$30 million a year, is derived from assets provided in 1974 by Daniel K. Ludwig, a reclusive US billionaire, whose fortune is derived from shipping interests.

Peter Newmark

NASA's old boot

Washington

THE National Aeronautics and Space Administration (NASA) announced this week that the problems uncovered by the failure of the space shuttle booster last month (see *Nature* 331, 2; 1988) should only delay the next flight by 6–10 weeks, moving the launch to the middle of August. While investigations continue into the failure NASA is planning to return to an earlier design of the offending part.

The carbon-coated boot which protects the movable joint where the nozzle is attached to the rocket had been altered between last August's test, which was successful, and December's, which was not. But in the August firing, the nozzle was gimbaled through an angle of only four degrees, compared to the seven-degree swivel applied in December, and it is therefore not clear whether the earlier design is actually any more robust. NASA now believes the major failure of the boot occurred after the test firing.

In the meantime, engineers at Morton-Thiokol, the manufacturer of the booster, have denied newspaper stories that an additional flaw was found after the December test. The nozzle assembly is sealed to the booster casing by three O-rings, and there were reports that damage had been discovered in this joint. But a spokesman for Morton-Thiokol says that although some hot gases got past a simple adhesive strip, the O-rings were not breached.

The demands being made of the re-designed booster are considerably more strenuous than those originally required before the first shuttle flew, and there has been speculation that the kind of problem that occurred in December's test might have happened during a previous flight without being noticed. David Lindley

UK threat to nature reserves

London

BRITISH conservationists have reacted with alarm to revelations this week that the government is to examine ways of selling state-owned National Nature Reserves (NNRs). The Nature Conservancy Council (NCC), a nominally independent state-funded body which maintains and manages NNRs and advises the government on conservation policy, has been asked to extend a routine review of its assets to include an investigation of the feasibility of selling NNRs. Mr Nicholas Ridley, Secretary of State at the Department of the Environment, which this year will provide the NCC with £36.5 million as grant-in-aid, has responded defiantly to the resulting public clamour.

Of the country's 241 NNRs, covering a total of 400,000 acres, 61 are owned outright by the NCC and 48 are part-owned. The remainder are in private hands. Since 1981, NNRs have been designated as sites of special scientific interest, precluding landowners from carrying out specified "potentially damaging operations" without informing the NCC.

Precisely who would want to buy an NNR remains unclear. Voluntary conservation bodies such as the Royal Society for the Protection of Birds and the 48 local Nature Conservation Trusts (coordinated by the Royal Society for Nature Conservation) are adamant that they could not afford to buy or maintain NNRs.

Within the NCC there seems to be little enthusiasm for the idea. The privately expressed fear is that being the paymaster, the government could eventually have the last word if, upon receiving the results of the current review, due to be completed by the summer, it decides that the NCC did not warm to its task sufficiently energetically.

Simon Hadlington

DoE provides funds for human genome sequencing

San Francisco

As the debate rages in Washington over what should be done about mapping and sequencing the human genome, who should do it and how it should be done, the US Department of Energy (DoE) has given three of its national laboratories a free hand — and some money (see p. 101) — for finding technologies to carry out the project.

Last year, DoE forced the pace by naming Lawrence Berkeley and Los Alamos National Laboratories as human genome centres. That move may have been premature, says Dave Smith, currently director of DoE's efforts, but work will not necessarily be limited to these facilities in the future.

Los Alamos, building on its GenBank experience in computing and data-handling, is studying methods for storage and distribution of the vast amounts of data to be generated by the project. Experiments are also under way on computer models of genome organization in order to predict the barriers to mapping and sequencing presented by the large amount of repetitive DNA in the human genome. A robotics project aims to automate some of the tedious jobs involved in DNA and sample preparation.

The choice of Lawrence Berkeley Laboratory as a centre for genome work came about partly because of its proximity to the University of California campus at Berkeley. Scientists at Lawrence Livermore Laboratory, some 30 miles away, were surprised that they too were not granted official status as a centre, but expect to be closely involved in the project.

Livermore has been in the field for five years, through its development of flow cytometry for sorting human chromosomes, and the resulting National Gene Library Project, which has generated small-insert phage libraries for each human chromosome and has now begun the next phase: large-insert phage and cosmid libraries that will be vital to the mapping effort.

All three laboratories are considering the newly available mapping approaches that use the rare-cutting restriction enzyme, *NotI* to generate DNA fragments of up to 1,000 kilobases in size, which can then be fractionated on pulsed-field gels. Scientists at Berkeley are concentrating their efforts on Yeast Artificial Chromosome (YAC) vectors, which have the capacity to accommodate the *NotI* fragments. Worries about difficulties in cloning fragments in YAC vectors led the Livermore team to attempt an alternative approach, in which the 100 or so *NotI* fragments per chromosome are ordered

through the use of 'linking probes' which span *NotI* cleavage sites. Specialized vectors for the selective cloning of such probes are being developed at Livermore.

Los Alamos and Livermore scientists have begun their mapping efforts in earnest, starting with chromosomes 19 at Livermore, and 16 at Los Alamos, both chosen for their size and concentration of clinically relevant markers, including the DNA repair genes, in which both laboratories have research interests.

The Livermore team is using computer modelling to optimize the DNA fingerprinting method, which determines clone overlap through a comparison of restriction patterns. Using fluorescently labelled DNA, enzymes that generate 80–100 fragments per cosmid, and an Applied Biosystems automated sequencer to read and store the data from the gels, project director Tony Carrano estimates that the team could map chromosome 19 to near-completion in 110 gel-days, a project they will begin in a month or so.

Los Alamos scientists have their eyes on the so-called 'random probes' method, proposed by Hans Lehrach, now at the Imperial Cancer Research Fund in London. This is a gel-free technique in which random oligomers are hybridized to arrays of thousands of DNA spots, or possibly even bacterial colonies, on filters. A computer program would then use the hybridization data to order the clones. Despite the attractiveness of a gel-free system that might not even require DNA preparation, the method is unproven, with many technical hurdles yet to surmount. In the meantime, the Los Alamos team has begun limited mapping of chromosome 16 by working with phage clones, starting with clones that have already been genetically mapped to sites on the chromosome.

Beyond the genome-mapping phase, which most expect to complete within 3–5 years, looms the arduous task of sequencing. Lawrence Berkeley scientists are counting on automated advances in current sequencing technology, such as those being developed by the Japanese, to bring the cost down tenfold from the current \$1 per nucleotide. Los Alamos scientists are gambling on an effort to develop gel-free DNA sequencing using a flow system in which a single fluorescently labelled DNA molecule would be bound to a matrix, progressively digested with endonuclease, and the bases identified by flow cytometry. Although early results are promising, the project is a long-shot, says Ed Hildebrand, coordinator of the genome effort at Los Alamos, but one worth risking.

Marcia Barinaga

Sex test banned

New Delhi

MAHARASHTRA has become the first state in India to ban prenatal sex determination testing on pregnant women. Chief Minister S.B. Chavan said the decision was taken because of "deep concern" over the widespread abuse of the test for aborting female fetuses. Legislation is to be introduced in February.

The ban applies only to private hospitals and laboratories. Government-run hospitals and research institutions of the Indian Council of Medical Research can carry out the test provided the woman is over 35 years of age and already has a child with an inherited defect.

Selective abortion of females has long been the subject of controversy in India and Maharashtra's action follows complaints from women's groups. Although gynaecologists are divided over the wisdom of a blanket ban on private practitioners, it is likely that more states will follow Maharashtra's.

K.S.J.

Heineken award

London

THOMAS R. Cech, of the University of Colorado, Boulder, is to receive the 1988 Dr H.P. Heineken Award for his discovery that RNA can act as an enzyme. The award, worth Dfl250,000 (\$130,000) is the largest Dutch prize in cash for scientific research. It is awarded every two years by the Heineken Foundation on the nomination of the Royal Netherlands Academy of Arts and Sciences. Cech's breakthrough, made in 1986, raises new questions about molecular evolution, with the possibility that RNA was the primordial prebiotic molecule and that proteins and DNA evolved subsequently.

S.L.H.

US research

Washington

THE latest figures compiled by the National Science Foundation (NSF) show that Johns Hopkins University is still the largest recipient of federal research dollars, largely on the strength of money allocated to the Applied Physics Laboratory for defence-related research. The NSF data are for the 1986 fiscal year. The government committed \$6,538 million for research and development to all US universities and colleges, of which 85.5 per cent went to the top 100 institutions, and 25.9 per cent went to the top ten.

	J.P. \$000
Johns Hopkins University	445,718
Massachusetts Institute of Technology	188,120
Stanford University	180,186
University of Washington	146,718
University of California, San Diego	133,243
Columbia University, main division	127,131
University of California, Los Angeles	125,483
University of Wisconsin, Madison	120,626
Cornell University	112,707
Yale University	111,687

French government outlines new university reforms

Paris

JUST over a year after a disastrous attempt at university reforms was aborted, the French government has decided to have a second try. An outline of the new proposals was published last week (7 January) by minister-delegate for research and higher education Jacques Valade, following the final report (*Demain l'Université*) of a 69-member commission (see *Nature* **328**, 102; 1987). Inevitably controversial — the historic Paris riots of 1968 were intimately linked to problems in the French system of education — the present proposals have so far provoked only scepticism.

Although Valade's remit also embraces research, his outline reforms apply only to university higher education. Unlike those elsewhere, French universities have a monopoly on neither higher education nor research. Basic research is almost entirely carried out in state research establishments with a separate funding structure and administration, themselves targeted for reform, while the elite *grandes écoles* (engineering and military academies) and private colleges cream off many of the brightest undergraduates in science and engineering.

The highest of Valade's four major priorities is an attack on the present 'paradoxical' university admissions procedure. In order to compete with other institutions of higher education, access to degree courses is selective even though universities have a statutory obligation to accept any candidate with a *baccalauréat* (examinations sat at the end of secondary education). Students not admitted to a degree course, or who choose less specialized studies, are automatically able to register for a two-year course leading to a national diploma of general higher education (DEUG).

But the DEUG has a high failure rate, attributed to high staff-student ratios and uneven student abilities. Attempts to introduce other short courses geared to vocational training or degree entrance have also been forced, by oversubscription, to specify higher admission requirements.

Valade feels that this 'selection by failure' is unacceptable and has adopted his advisory committee's proposal for a new kind of 'university college' that offers short courses leading to national diplomas or preparation for degree-level studies. These colleges, which would be governed and financed by the universities, although geographically separate, should, Valade feels, be "better adapted" to students' abilities and to the "rapidly changing employment opportunities" available.

Although the opposition socialist party also recognizes the need to tackle the 'first cycle' (first two years) of university education, the idea of university colleges has already raised the hackles of left-wing lecturers' unions. They feel that the scheme is no more than a watered-down version of the two-tier structure proposed by Valade's predecessor, Alain Devaquet, which led to student riots and his resignation in December 1986 (see *Nature* **324**, 501; 1986).

Valade's proposals, which also provide for greater autonomy of university gov-

erning bodies, come just a few weeks after his colleague, education minister René Morony, announced plans for a FF28,000 million (£2,800 million) overhaul of primary and secondary education. While the cost of Morony's plan has raised eyebrows in the Ministry of Finance, the absence of figures for Valade's scheme has led to incredulity — particularly with respect to plans to increase student grants and to raise student intake by the year 2000.

The French press and government critics are saying that the proposed reforms are no more than an attempt to enter May's presidential elections with an education policy based on what Valade has called a "consensus" — the "independent and impartial" report of his advisory committee.

Peter Coles

Universities defy South African government

Cape Town

IN South Africa, the academic year has drawn to a close without the government enforcing the regulations promulgated on 16 October (see *Nature* **330**, 4; 1987) empowering it to cut subsidies in the face of political activity on university campuses.

The most radical act of defiance of the regulations so far has come from the University of the Western Cape (UWC). Its students responded with a week-long boycott of lectures, an incident which should, in terms of the regulations, have been reported to the government with a notification of what action was being taken against offenders. Instead, the university administration simply notified the Minister of Education and Culture for Coloured Affairs, Mr Carter Ebrahim, that it did not intend reporting on the boycott or taking any action.

The administration of the regulations is complicated by the bizarre fact that there are no fewer than five ministers responsible for education in South Africa. The regulations are the brainchild of Minister of National Education F.W. de Klerk, who assumes a coordinating role and is also front-runner in the stakes to succeed P.W. Botha as state president. Responsibility for the implementation of the regulations, however, lies with the minister in whose department the particular university in question falls. In the case of the 'white' universities, four of which are open to all races, this is white education minister Piet Clase, whereas the ethnic universities (one for each African tribal group) are the responsibility of the minister for African education, Dr Gerrit Viljoen, himself a former principal of the Rand Afrikaans University. UWC (also now open) falls under the aegis of coloured affairs minister Ebrahim and the University of Durban-Westville falls under Indian education minister K. Ramduth.

These ministers do not appear to have entirely congruent views on the regulations. Ebrahim has already omitted to act in the case of UWC. When the rector of the University of Zululand, Professor A.C. Nkabinde, objected to Viljoen about "the whole concept of a university being policed by an education authority", the minister reportedly agreed with him and assured him that his university would not have the regulations applied.

Three universities are applying to the Supreme Court to have the regulations declared invalid on the grounds that they are both *ultra vires* and so vague and loosely worded as to make their implementation impossible. Papers have already been filed in the Supreme Court in Pietermaritzburg by the University of Natal, and the case is expected to be heard on 9 February. In the same week, the Cape Town Supreme Court will hear separate applications by the Universities of Cape Town and the Western Cape. Rhodes University and the University of the Witwatersrand have decided not to proceed with court proceedings, apparently on the grounds that their counsel is not of the opinion that legal action is likely to be successful.

It is rumoured that the chancellors of the universities that have opposed the regulations will be petitioning the government soon, in an attempt to prevent the regulations being implemented once the academic year starts again in mid-February. Such a deputation is likely to include two of the most powerful men in South African industry: Harry Oppenheimer, retired chairman of Anglo-American Corporation (South Africa's largest mining house) and chancellor of the University of Cape Town, and Mike Rosholt, chancellor of the University of the Witwatersrand and chief executive of Barlow Rand, the country's largest manufacturing company.

Michael Cherry

Japan's new international superconductivity centre opens

Tokyo

JAPAN'S International Superconductivity Technology Center, set up with massive funds from the private sector at the urging of the Ministry of International Trade and Industry (MITI) opens today (14 January). Foreign (that is, non-Japanese) companies are welcome to join — but at a price.

A committee of academics and industrialists formed by MITI in response to the explosion of interest in the new high-temperature superconductors recommended the establishment of the centre in a report released last September. The centre will both gather and disseminate information and establish a laboratory for research (see *Nature*, 330, 2; 1987).

More than fifty companies, including electric utility, electronic and cable manufacturing companies, mining concerns, a gas corporation, steel and glass producers, shipbuilders, car manufacturers, even spark-plug producers and banks have signed up to join the information centre, membership of which requires a down-payment of ¥2 million (about \$16,000) plus an annual subscription of the same amount.

The centre will hold international symposia, seminars and workshops, gather and disseminate information in a newsletter, invite foreign researchers to give lectures in Japan and also carry out public education campaigns.

But the budget for the information centre is nothing compared with the funds going to the research laboratory. About 40 companies will donate ¥100 million each to set up the laboratory and ¥12 million a year for running costs, over and above the donations for information. This brings the total budget for the centre to about ¥4,700 million (nearly \$40 million) in its first year of operation.

The laboratory will open in October this year on a site in Tokyo rented from the Tokyo Gas Corporation, one of the founding members. There will be five research laboratories covering oxide and non-oxide research, processing technology, physical property evaluation and research into the theory. A research laboratory in Nagoya will also carry out research on ceramics.

The companies that join the laboratory can send one or two researchers. And they can also have members on the board, the decision-making body of the centre. Rights to patents resulting from research will be shared by the centre and the company of the researcher that makes the discovery. And, according to MITI officials, foreign companies are welcome to join, provided they pay the fee. The centre may also

employ its own researchers, including foreigners, in addition to those from the companies.

MITI is also studying the possibility of inviting researchers to the laboratory from foreign public research institutions. But, realizing that the membership fee may be beyond the reach of such organizations, MITI will adopt a 'flexible' approach when determining the participation fees when the need arises.

MITI has already contacted several US and European companies, and brochures in English have been sent to the US Chamber of Commerce. MITI is now hunting for a suitable European organization to which to send pamphlets. But, according to Risaburo Nezu, of MITI's Agency of Industrial Science and Technology, it is difficult to decide whom to send pamphlets to as "there are so many countries". Suggestions would be 'welcome', he says.

MITI is clearly keen to make the centre international to help diffuse US criticisms that Japan has not been making a fair contribution to international basic research

(a criticism which even Japanese government officials widely accept). And while some in the United States and Europe may suspect MITI's intentions (is it another plot to steal ideas from the West?), the ministry is obviously striving to establish a truly international research centre. And, judging from the rapid rate of developments in this field in Japan (for

Apparently it can keep ¥4,700 m
Circulating for up to a year...



example, see article below) foreign countries may be well advised to take advantage of the opportunity to join.

David Swinbanks

Another breakthrough for Japan in superconductor technology

Tokyo

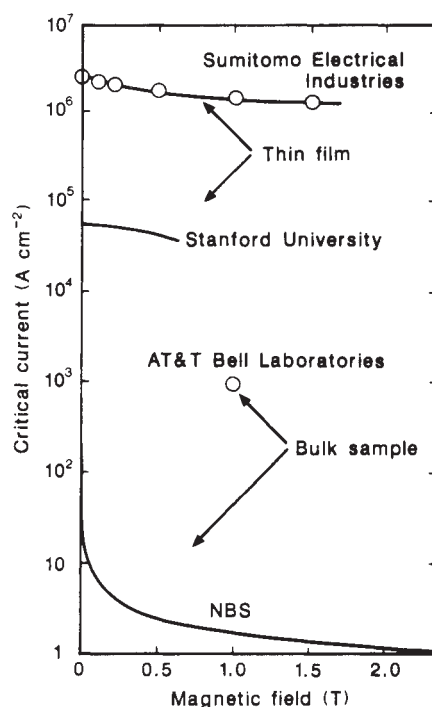
JUST when all seemed quiet on the high-temperature superconductor front, Japan has announced another 'breakthrough'. Researchers at Sumitomo Electric Indus-

tries have developed a single-crystal thin film that can carry a maximum current of 2.5 million A cm⁻² at the temperature of liquid nitrogen; high current-carrying capacity is also maintained in a strong magnetic field.

The development, announced last week (6 January) breaks the previous record set by Nippon Telegraph and Telephone (NTT) (1.8 million A cm⁻² in July last week (*Nature* 329, 98; 1987), and the technique to make the film differs from that of NTT.

Holmium-barium-copper oxide was sputtered epitaxially on the <100> crystal face of a monocrystal MgO substrate (as opposed to the <110> face of a SrTiO₃ substrate in the case of NTT). As a result, the *a* and *b* crystal axes of the ceramic lie perpendicular and parallel respectively, to the plane of the film and high current-carrying capacity is thus possible in both the vertical and horizontal directions in the film.

One major problem besetting the development of the new high-*T_c* superconductors has been their inability to carry high currents under a magnetic field. But Sumitomo's thin film maintains a critical current density of 1.5 million A cm⁻² under a magnetic field of 1 tesla. And the current-carrying capacity of the film has remained stable for over a month. David Swinbanks



Scandal rocks nuclear power industry in West Germany

Munich

A BRIBERY scandal linked to the illegal shipment and storage of nuclear waste has dealt a serious blow to the image of West Germany's nuclear power industry. And after a year's lull, the question of long-term storage sites for nuclear waste has once again taken centre stage.

West Germany's centre-right coalition government seemed to have outlasted the intense anti-nuclear sentiment generated by the Chernobyl catastrophe (see *Nature* 321, 640; 1986). But the current scandal has provided ammunition to those in West Germany who believe that it is not possible to run a safe nuclear programme.

The scandal began in early December

when bribery charges were brought against employees of the nuclear waste transport firm Transnuklear, located in the city of Hanau. One of the employees committed suicide in prison after admitting his guilt.

The affair began to loom larger when, on 17 December, it was revealed that Transnuklear had shipped more than 350 barrels of radioactive waste under false labels to temporary storage sites in West Germany. The waste was intended to be processed in a Belgian nuclear energy centre and returned.

Mislabelled barrels began turning up all over the country and the numbers grew day after day throughout the Christmas

holidays. The Environment Ministry's latest count has reached 1,900 barrels. It was also discovered that 321 of the barrels were contaminated with traces of plutonium and cobalt-60.

Environment Minister Klaus Töpfer (Christian Democrat or CDU) has been hard-pressed to keep up with the revelations. A 'working group' of representatives of the Environment Ministry and the *Länder* (states) last week recommended stricter regulation of the nuclear waste transport industry. Most urgent is the need to transfer the oversight of the industry to the federal level; four different authorities at present regulate the transport of various radioactive materials.

Chancellor Helmut Kohl (CDU) recognized the need for damage control on 7 January when he criticized Transnuklear employees for their improper behaviour. But this attempt to limit the political consequences of the scandal seems to be too little, too late. There are reports of a considerable loss of public trust in the government and the nuclear industry. Green Party Bundestag member Barbara Rust said that public disapproval of the continued reliance on nuclear power — measured at up to 70 per cent according to independent surveys in late 1987 — "can only go higher". The latest focus is on so-called 'permanent storage' (*Endlager*) for nuclear waste of all types. Nowhere in the world is there a site that opponents of nuclear power think secure enough for the storage of used nuclear fuel elements, which contain plutonium and other highly radioactive substances.

Construction on the proposed West German site, located at the former salt mine of Gorleben, was halted after a fatal accident in May last year and will not begin again until the spring.

If it should not prove feasible to complete the Gorleben site, the government announced on 7 January that it would eventually propose another site. Werner Herud of the federal agency responsible for preparing waste storage sites, the Physikalische Technische Bundesanstalt, said that all the alternatives are located in nearby salt beds. But some scientists have long argued that salt is an inappropriate medium for waste storage. Controversy has also arisen about the proposed storage site for low- and middle-radioactive waste, known as Schacht Konrad, which is also based on salt.

Even if all West Germany's nuclear power plants were shut off immediately, there would still be a problem of what to do with the waste. Until now, spent West German fuel elements have been stored at the French facility at Cap La Hague; beginning in 1993, however, they are scheduled to return. The question is, will there be some place to put them?

Steven Dickman

New AIDS committee in France — the buck stops here?

Paris

To stem what is perceived as a proliferation of competing organizations, committees and charities concerned with AIDS (acquired immune deficiency syndrome), the Institut Pasteur and the Fondation pour la Recherche Médicale (FRM) have announced yet another — the Comité National pour la Recherche sur le SIDA.

The reasoning behind this seemingly paradoxical move, according to Mr CoursMach of the FRM, is "to ensure the best possible national coordination for distributing donations from the public, while providing donors with an assurance that the money will be well-spent". The committee, whose executive council is drawn equally from the Pasteur and FRM's scientific advisory council, will select applications for funding "on the basis of scientific or humanitarian merit" and will, through its membership, ensure coordination with state-supported research.

The association with the Pasteur Institute is also a tactical step towards the creation of a National AIDS Research Foundation. The new foundation should provide a more visible target for fund-raising and, according to CoursMach, "ensure the centralization of funds and avoid a dispersion of donations".

The FRM, France's largest medical charity, was chosen by the government as the focus for a national fund-raising drive last June and has already distributed FF10 million (£1 million) to research groups — including the Pasteur itself — and to smaller voluntary bodies.

Coming so soon after the creation of two other AIDS foundations associated with the Pasteur Institute — the French and American AIDS Foundation and the

World AIDS Foundation (see *Nature* 330, 507; 1987) — the announcement of the new committee caused some initial confusion. "The idea", explained Dr Caroline Chaine of the Pasteur Institute, "is to rationalize the work of different associations, whether private or state-funded, and help them to be more effective, while respecting their independence". "The number of AIDS foundations and committees is confusing", admitted Chaine, "but this should be the last one for quite some time."

Peter Coles

Paris

● LATEST World Health Organisation (WHO) figures suggest that the number of known cases of AIDS in France increased from 1,980 at the end of June 1987 to 2,523 at the end of September — an increase of 25 per cent. This means that France has the second highest proportion of declared AIDS sufferers in the world, after the United States, which had 45,436 known cases on 9 November. Britain had 1,123 cases.

Danielle Mitterrand, wife of the French President, and founder of France-Libertés, a charitable organization aimed at promoting civil and human rights, recently called attention to the cultural and financial barriers to AIDS prevention in some African countries. In a letter to the daily newspaper, *Le Monde*, Mme Mitterrand pointed to the inadequacy of resources for screening blood supplies for AIDS antibodies and severe obstacles, both financial and logistical, to effective educational campaigns.

In the same issue of *Le Monde*, it was pointed out that a pack of condoms cost the equivalent of a day's salary in some parts of East and Central Africa. □

Another wrinkle in patchwork of US environmental release

Washington

YET another layer of complexity will be added to the current regulatory patchwork governing the environmental release of recombinant organisms if plans at the US Environmental Protection Agency (EPA) come to fruition. Last week, the EPA's biotechnology science advisory committee voted to go forward with a scheme to institute a network of review committees — to be named 'environmental biosafety committees' — for the purpose of approving field tests of genetically altered organisms.

The issues of what constitutes an environmental release and at what point research involving recombinant organisms falls under the EPA's regulations have been murky. The incident last summer when Montana State University researcher Gary Strobel injected elm trees with a recombinant bacteria without EPA approval (see *Nature* 328, 659 & 329, 95; 1987) emphasized the fact that these issues required clarification. EPA is now in the process of extending its Toxic Substances Control Act to cover research and development work, and the environmental biosafety committees would oversee the application of this statute in biotechnology research settings.

EPA plans to model its environmental biosafety committees on the institutional biosafety committees, administered by the US National Institutes of Health (NIH) Recombinant Advisory Committee, that review all recombinant DNA experiments. There is wide agreement that the institutional biosafety committees have been a good way to control recombinant DNA experimentation without

being unduly restrictive. As outlined, the EPA plan calls for setting up committees of five people — three scientists with expertise in areas of microbial or plant ecology and two representatives from the local community — at each university or company that would field-test recombinant organisms. The EPA's committees would differ from the NIH biosafety committees in that they would be backed up by regulatory statutes and infringers could be prosecuted.

The case-by-case review of proposed experiments by the environmental biosafety committees would also include the solicitation of public comment on the field test, and the committee would be responsible for addressing questions from the local community. The survey of attitudes toward biotechnology sponsored by Congress last year (see *Nature* 327, 453; 1987) showed that the public had confidence in university scientists' assessment of environmental risks. But the survey was completed before the Gary Strobel 'incident', and it is not clear whether that confidence still holds.

The environmental biosafety committee concept is likely to stir opposition from industry and university researchers, who are already confused by the federal maze of regulations governing biotechnology. Henry Miller, the biotechnology adviser to the US Food and Drug Administration, says EPA's plan to set up environmental biosafety committees "puts the cart before the horse", because the pivotal issue is deciding which experiments should be subject to review. Once this is ironed out, he reasons, there is no need for the additional bureaucracy that the review committees would certainly add.

William Gartland, the director of the recombinant DNA activities office at NIH, wonders whether the NIH institutional biosafety committees already in place would not be sufficient to review field release experiments, particularly at institutions already heavily involved in recombinant work. The chairman of the institutional biosafety committee at Montana State University, Clifford Bond, says he would "hate to see more committees" because it is the "worst way to get anything accomplished". But he concedes that it "certainly helps to have a hammer" of firm regulations to drop on those who would break the rules.

EPA is devoting increased staff time to working out the details of how the environmental biosafety committees would function, and the rules for establishing the committees is expected in the spring.

Carol Ezzell

Health risks of radon are given a new look

Washington

How much does environmental exposure to radon and its alpha-emitting daughters contribute to lung cancer in humans? This question has received renewed attention recently with the realization that radon exposure can be relatively high in homes in certain parts of the United States. In a report released last week, a committee of the National Research Council proposes new estimates of the risk, concluding that the effects of radon exposure vary not only with dose, but also with time since exposure and with age.

To reach the new estimates, the research council's committee on the biological effects of ionizing radiation used statistical regression techniques to evaluate data from four studies of radon-exposed miners. The committee found that risk of lung cancer attributable to radon reduces with age. The analysis also revealed that exposure within the previous fifteen years was a better predictor of risk than any earlier exposure. Cigarette smoking appeared to interact with radon exposure in a multiplicative rather than additive manner in increasing the risk of lung cancer. The committee urged therefore that more data be collected on the smoking habits of miners exposed to radon.

In evaluating cumulative radon exposure, the committee used the working level month (WLM) which is defined as exposure to approximately 100 pCi of radon per litre of air in 170 hours, a time period chosen for historical reasons. Occupational exposure to 4 WLM, approximately the number of hours in one month, would increase lung cancer risk by 60 per cent for the general population for people between 20 and 40 years old. The greatest increased risk is to smokers, who are at ten times higher risk than non-smokers.

Polonium, radium, thorium and uranium are also alpha-emitters presenting health risks, as are the transuranic elements. Cancer risk from environmental exposure to the naturally occurring elements is relatively low. Epidemiological studies of exposure to the transuranic elements does not provide sufficient data for risk estimates.

Joseph Palca

Health risks of radon and other internally deposited alpha-emitters. (National Academy Press, Washington, DC 1988.)

AIDS Commission: Correction

THE chairman of the US Presidential Commission on the HIV epidemic is retired admiral James D. Watkins, not James D. Watson (see *Nature* 330, 687; 1987). □

Dutch field trials

Waalre, The Netherlands

FIELD trials of genetically modified plants are expected to be carried out for the first time in the Netherlands this year. The biotechnology company Mogen International has succeeded in inserting a foreign gene into potato plants conferring resistance to viral infection, although the identity of the gene and the disease have not been disclosed. The *ad hoc* Committee on Recombinant DNA has recommended approval of field trials, and official permission from the community of Dronten, where the trials are to be carried out, is expected to be a formality.

Permanent legal rules for field tests with genetically modified organisms are expected to be incorporated in the law for environmentally dangerous substances later in 1988.

Casper Schuurin

Protocols for Turin shroud

SIR—One of the essential points of my previous remarks about the confidentiality surrounding the dating protocols for the Turin shroud (*Nature* 327, 10; 1987) is that the matter unavoidably involves religious passions. No better demonstration could be found than P.R. Smith's denunciation (*Nature* 328, 11; 1987) of my letter as a "gross insult" to the experts whose job it has been to design the procedures for the carbon tests. Smith suggests I am hostile to religion; but anyone who has been critical of attempts in recent years to give scientific legitimacy to the shroud will be accustomed to such accusations.

The point is, however, that there has been unacceptable secrecy and confusion about how the tests are to be conducted, beginning with an early report in *La Stampa* (Turin, 5 October 1986), at the conclusion of the closed conference of experts, indicating that the "timetable and methods of investigation are secret". In this respect, the open response from Harry Gove (*Nature*, 327, 652; 1987) is welcome, but it stands in stark contrast to a 'Vatican spokesman' whose reaction to my previous letter in *Nature* confirmed the suspicions it expressed. According to a wire service report, the spokesman told the *London Daily Telegraph* that it was indeed likely that "only the results" of the tests would be made available to scrutiny by independent observers — precisely the issue that is so unsettling — adding that they "would be made available in a couple of years". This last assertion also curiously contradicts the initial announcement about the tests, which stated that the results were to be made public at Easter 1988.

I hope and expect that there will be a detailed and satisfactory disclosure of the test protocols before the samples are taken. An important first step would be for the person who is in a position to speak authoritatively about the conduct of the tests to identify himself or herself. I also trust that M.S. Tite is correct in suggesting (*Nature* 327, 456; 1987) that neither the British Museum nor the seven testing laboratories will be party to the tests unless each of them is confident that the protocols absolutely preclude tampering with the samples by the introduction into the chain of evidence of ¹⁴C-depleted linen, such as mummy linen.

There is no specific reason to believe that any of the scientific devotees of the shroud or its Vatican owners would seek to rig the tests. Nevertheless, besides the obvious involvement of religious sensibilities, all of us must face the fact that a veritable industry has been built up around the shroud. In this respect, the Vatican can rightly be seen as having a vested interest in keeping alive at least the

possibility that the shroud is the actual burial cloth of Jesus. As a negative test result is clearly going to spoil the fun, it is imperative that the protocols for the sample handling be spelled out in advance and that they be seen to be beyond reproach.

DENIS DUTTON

*School of Fine Arts,
University of Canterbury,
Christchurch, New Zealand*

Our daily bread . . .

SIR—The Committee on Medical Aspects of Food Policy of the Department of Health and Social Security has set up a panel to review the "recommended daily amounts" of food energy and nutrients for groups of people in the United Kingdom. The panel is concerned that its deliberations should encompass all opinions in this field and invites concise written submissions based on reasoned argument and scientific data from interested parties.

M.J. WISEMAN

*Nutrition Unit, Room C307,
Department of Health and Social Security,
Alexander Fleming House,
Elephant and Castle,
London SE1 6BY, UK*

Embryos sacrosanct

SIR—In your leading article on the latest Instruction from the Congregation of the Faith (*Nature* 326, 229; 1987), several points are disputable. Not least, to subject science to human dignity does not mean ceasing to study developmental processes, but only to make more efforts to that end and to prevent accidents from happening. Not all that is technically possible is ethically admissible.

The instruction assesses human dignity for every human embryo from the very first stage of fertilized egg. Therefore, I do not see the need of any further determination of a 'time-limit' after which to *qualify* a zygote as a human embryo, especially by performing experimental tests on living embryos. Nor do I believe that the sorrowful destruction of human beings implied by today's techniques related to *in vitro* fertilization (IVF, a clear case of procured abortion) should be undervalued.

Given the moral illicity of heterologous IVF and artificial insemination, the explanation of the moral illicity of heterologous IVF and artificial insemination (except for a few defined exceptions to the latter) seemed to me clear and full of details. The main point is that even using gametic cells coming from a married couple, the possibility of generating a child depends on the technical competence of other people outside the family.

The deep meaning of the conjugal act is lost.

Political and legislative authorities, therefore, ought to prevent any kind of experimentation on human embryos, including the development of IVF technique, in the knowledge that no true progress may be achieved through the denial of human dignity since its first appearance. I hope that the day will never arrive when a man will choose to kill another human being to save an animal, but this depends on our actions today.

ANGELA M.S. LEZZA

*Emory School of Medicine,
Department of Biochemistry,
Woodruff Memorial Building,
Atlanta, Georgia 30322, USA*

Tonegawa's prize

SIR—It is difficult to understand the assertion in "Nobel prize for Japanese immunologist" by Peter Newmark (*Nature* 239, 570; 1987) that a contributory factor in Susumu Tonegawa's being able to dominate his area of research was "others . . . [being] . . . more constrained than Tonegawa by the then current problems of carrying out recombinant DNA research". We would like to call your readers' attention to two points:

(1) The paper cited by Newmark (Hozumi, N. and Tonegawa, S., *Proc. natn. Acad. Sci. U.S.A.* 73, 3628; 1976) reports no experiments with recombinant DNA.

(2) Recombinant DNA work in Switzerland was overseen by a committee of the Swiss Academy of Medicine, which promulgated guidelines as strict as those anywhere in the world, inspected facilities and registered experiments. To the best of our knowledge, neither Tonegawa nor anyone else at the institute ever carried out an experiment that would have been impermissible at the time in the United States or the United Kingdom.

Surely, the pressures of an editorial deadline were responsible for the omission of the *locus operandi*. Tonegawa's Nobel prizewinning work was done in Switzerland at the Basel Institute of Immunology. The present and former members of the institute, as well as our sponsors, F. Hoffmann-La Roche & Co., are justifiably proud of Tonegawa's achievements and would like to believe that the research environment of the institute was a real contributory factor in his success.

CHARLES STEINBERG, LOUIS DU PASQUIER, KLAUS KARJALAINEN, HANSRUEDI KIEFER, ZDENEK TRNKA, POLLY MATZINGER, UNA CHEN, GHOLAMREZA DASTOORNIKOO, ANDRE TRAUNECKER, YASUSHI UEMATSU, ANDREW KELUS, WERNER HAAS, LUCIANA FORNI, JOSEPH SCHWAGER, PAWEŁ KISIELOW.

*Basel Institute for Immunology,
Grenzacherstrasse 487,
Postfach, CH-4005 Basel, Switzerland*

Ghost-like neutral particle observed

The positrons of distinctive energy observed in heavy-ion collisions over the past three years have now been mirrored by the surprising discovery of pairs of photons in similar circumstances.

THREE years have passed since people studying heavy-ion collisions were first struck by the observation of positrons with a characteristic energy of just above 300 keV among the collision products, but nobody is yet very much the wiser. It now seems generally to be agreed that the purported explanation of the positrons as a kind of incidental consequence of the disturbance of the Dirac sea of electrons of negative energy by the potential field of two heavy nuclei in collision (see *Nature* **320**, 209; 1986) will not by itself suffice. Indeed, with the passage of time, evidence has further accumulated to support the original claim that there is a new phenomenon to be explained, but a definitive explanation seems still a long way off.

Part of the trouble now is that there is almost too much observational evidence to be accounted for. At the West German heavy-ion laboratory at Darmstadt, experiments have now demonstrated both that the positrons observed form pairs with electrons of equal energy (which is not surprising) but also that there are no fewer than three characteristic energies (624, 720 and 815 keV for the pair of electrons) in collisions between uranium and thorium ions (carrying just under 6 MeV of energy per nucleon). More recently, a French group working at Darmstadt has found evidence for a fourth energy-level, at just under 1,500 keV.

But now there is even bigger surprise. A group from Stanford University and the Lawrence Berkeley Laboratory working at the Lawrence Berkeley ion accelerator called super-HILAC have found pairs of oppositely travelling photons among the products of the collision of uranium and thorium ions at the same energy of 5.9 MeV per nucleon (K. Danzmann, W.E. Meyerhof *et al. Phys. Rev. Lett.* **59**, 1885; 1987). By all accounts, the development has given super-HILAC's supporters, recently dismayed that the average energy per nucleon is so much smaller than likely to be available at the latest generation of ion-collision machines, reason to hope that there will be a new lease of life for it.

From the outset, it has been apparent that one way of accounting for the original observations, and perhaps the most intriguing, would be to postulate the existence of an electrically neutral particle capable of spontaneous decay into an electron pair. Indeed, again since the beginning, there has been talk of the possibility that the neutral particle concerned

might be the elusive axion — the particle whose existence is suggested by the need to reconcile charge and parity non-conservation in weak interactions with the observation that strong nuclear interaction behave as people would expect, but which has also been widely canvassed as the origin of the missing mass (over and above that observed in galaxies) required to close the Universe. But people have plainly been anxious not to make too much of that almost out of superstitious fear that to do so might make the phenomenon go away.

But if a neutral particle is capable of decaying into an electron-pair, may it not also decay into a pair of photons? That seems to have been the starting-point for the observations now reported from the work with super-HILAC. A little reflection will show that the detection of oppositely moving pairs of photons in an experiment in which an ion beam (U^{40+}) collides with a fixed Th target are far from child's play, if only because the frame of reference in which the emitted photons (both γ -rays) have equal energy is itself moving at a measurable fraction of the velocity of light (making necessary a 6° offset between the forward and backward directions of paired detectors. And then, given that the super-HILAC arrangement is bound to be an effective way of making X-rays and γ -rays copiously, there will be a huge background within which to search for coincident pairs of photons.

Evidently it should help to know at what energy to look for pairs of coincident photons, which is another of the reasons why the result of Danzmann *et al.* is such a surprise. For it turns out that there is indeed a statistically significant signal — a clutch of oppositely directed photons of equal energy — but that their mutual energy lies at none of the four known levels of the supposed neutral system recognized from the positron observations, but at a fifth level, corresponding to 1,062 keV (with a standard error of merely 0.1 per cent). The authors reckon they can exclude other decays into a pair of photons from states whose total energy is less than 2,000 keV.

What can be the origin of these puzzling signals? Modestly, the super-HILAC group acknowledges that its search for photon-pairs has been a hunt for needles in a haystack. In a range of energy in which the background produces nearly 700 coincidences of oppositely travelling,

equally energetic electrons, they have what they call an excursion of 167 events to hold up as their prize — the equivalent of five or six standard deviations, corresponding to a chance fluctuation of one in 10 million or so. In reality, the significance of their signal rests on the narrowness of its energy distribution; the 0.1 per cent line-width is taken to be a measure of the intrinsic properties of the state from which it is derived.

What kind of neutral particle may, when it decays, yield observable products with at least five sharply separated energies? The temptation to say that the truth may be more complicated in that there may be neutrinos as well, carrying away unmeasured amounts of energy, must be suppressed. The snag then is that the measured energy of the photon and electron pairs would not be sharp.

If, on the other hand, the unidentified neutral source of both the electron and photon pairs is a single particle such as an axion, it would be necessary to add the mass of an electron-pair to the kinetic energy of the electron-pairs observed at Darmstadt to obtain the allowed masses of the neutral particle. On that arithmetic, the state of the system observed at Berkeley is the lowest of those recognized. The supposed neutral system is thus a curious animal which, in its lowest state, decays into a pair of γ -rays and which has four more energetic states each of which decays predominantly into an electron-pair. There is no reason why an axion should satisfy these conditions. It is no wonder that people have begun to brood about the possibility that the structure concerned is some kind of atom constituted exclusively of fermionic matter. But ordinary positronium (an electron and positron bound together) will apparently not fill the bill.

Could it be that para-positronium, in which the two electron spins are parallel to each other, will meet the need, providing the long lifetime for the ground state implied by the narrow linewidth of the ground state? Whatever the explanation, it will also have to be explained why such an exotic material should be formed in heavy-ion collisions. Is it simply conjured out of the electrodynamic vacuum?

Where this surprising set of observations will lead is still, apparently, an open question. As always, the best resolution of the problem will be that people should collect more data. **John Maddox**

Archaeology

Old China for new

Paul G. Bahn

CHINESE archaeology has come to be characterized in the public mind by the terracotta army, thousands of lifesize troops buried in serried ranks guarding the tumulus of the emperor Qin Shihuangdi (third century BC) near Xian. Some of these figures have even become roving ambassadors of Chinese culture, and can be seen at an exhibition in London (see figure). This discovery follows in the tradition of the jade burial-suits and other art treasures seen in several countries in the 1970s. Spectacular finds are commonplace in Chinese archaeology, and since 1974, when the terracotta army was found by peasants drilling for water, there have been important discoveries in different regions which alter profoundly our view of the timing and scale of cultural development in Chinese prehistory.

One breakthrough occurred in 1983 when, at Niuheliang in the north-east of Liaoning Province (North China), the excavation began of six groups of stone tombs and a temple more than 200 m wide; a huge altar was uncovered by torrential rains about 50 km east of the temple. The temple is thought to be that of a goddess, as two small female figurines were found nearby; there are also elaborately designed jade ornaments, such as a carved dragon and, most remarkably, a 23-cm pottery head of a 'goddess' with her face painted red and inlaid eyes of green jade. Fragments of what may be her body were also recovered, and small female figurines were found at the altar.

All these finds are thought to be about 5,000 years old (Middle Neolithic), and have been attributed to the Hongshan culture. They represent the earliest archaeological evidence for any social elaboration, predating by at least 1,000 years the legendary Xia Dynasty (twenty-first to sixteenth centuries BC), previously considered to be the first Chinese civilization. The scale of construction shows a high degree of organization, and work is now in progress to open a further five tombs and to search for the remains of a city, some of which may lie beneath a huge square behind the temple.

Meanwhile, several important early cities have already come to light: a site was found in 1983 at Shixianggou, for example, 19 miles east of Luoyang on a fertile plain between the Luoshui and Yellow rivers. Excavation work since 1984 has uncovered a huge enclosure, 1.7 by 1.2 km, surrounded by a rammed-earth wall, 17 m thick, with at least seven gates. Inside are the remains of broad streets and a central compound comprising several

timber halls and palaces represented only by post-holes; these pits had timbers set on stone slabs inside them and were then packed with earth.

Excavations recovered arrowheads of jade and bronze, glazed pottery and inscribed 'oracle' bones, the earliest Chinese texts. Radiocarbon dates of older than 1500 BC from the palace confirm the styles of the bronze artefacts. This fact, together with historical sources mentioning an early capital at Shixiang, suggests strongly that this site is ancient Xibo, the lost capital of the Shang Dynasty of the seventeenth/sixteenth centuries BC.

There is a site of several square kilometres about 50 km west of Chengdu, the modern capital of Sichuan Province, where excavation was begun in 1929 by D. Graffin, a US archaeologist. Work resumed there recently, and in 1986 two big pits were found by workers at the local brick factory. These pits contained more than 400 and 500 objects, respectively, including some unique and remarkable bronzes, such as a mask 1 m in height with protruding eyes and ears, statues, and huge heads including one which seems to represent a bearded, western man. According to Professor Su Bingqi, president of the Archaeological Society of

China, this site could be the capital of the legendary Shu state, which disappeared in the late second millennium BC and is scarcely mentioned in Chinese history. It is likely that the remains of a huge palace lie beneath the modern city of Chengdu.

Yet another walled capital — this time that of the ancient warrior state of Qin, in Shaanxi Province — was found in 1985, and turned out to be one of the biggest and best-preserved sites in China. This was one of the many rural Chinese states, and its harsh political system and military skills enabled it to conquer and unify the Chinese world in 221 BC. It is in some ways analogous to Rome, and certainly its palaces and 13 royal mausoleums display architecture that rivals that of the classical world.

Finally, a very different discovery was made last year when a large box of white jade was found hidden in the cave of Lei Yin on Shijing Mountain, 45 miles from Beijing. It contained four progressively smaller boxes of stone and precious metals, and inside the smallest were two tiny cream-coloured pellets of bone. The cave is near the Yunju Temple, a historic Buddhist site, and it is probable that these pellets are fragments of Buddha himself. When he died in India, 2,470 years ago, Buddha was cremated, and the fragments distributed as sacred relics among his followers. It is known that 53 fragments were carried to China, three of which were given to a temple near Beijing in AD 616; and that the mother of the seventeenth century emperor Wan Li removed one



The emperor's terracotta warriors and horses from the Qin Dynasty, part of an exhibition at the Old Hall, Royal Horticultural Society, Vincent Square, London SW1, UK until 20 February 1988. The hall has been transformed into a replica of the emperor Qin Shihuangdi's tomb, in which 7,500 life-size terracotta warriors and horses were buried 2,000 years ago. The exhibition also contains many ancillary items from the first imperial dynasty. Information about opening times can be obtained by telephoning (0)1-828 2768. □

pellet, replacing it with two pearls. The two pellets found in the Leiying box were accompanied by two pearls.

Only the biggest and most spectacular finds in Chinese archaeology are reported in the west; the lesser-known, though equally important, data from more humble sites take far longer to emerge in syntheses by those few scholars in the west who can obtain and read Chinese journals. It is to be hoped that, as Chinese archaeology develops, some means will appear of keeping foreigners aware of new information rapidly and effectively.

The recent emphasis on artistic treasures may have helped to stimulate the flow of tourists to Beijing and Xian, but it has also attracted the greedy and unscrupulous: it is estimated that more

than 15,000 archaeological sites are already known in this huge country, and these are certainly a minute fraction of what exists. Even so, most of them remain unexplored through lack of manpower, resources and space to deal with the material. Consequently, gangs of thieves are operating intensively, digging up hundreds of ancient tombs and smuggling out thousands of artefacts, mostly via Hong Kong. Some were caught and prosecuted last summer, but the situation will certainly worsen, as in so many other countries, unless the market in smuggled antiquities is stamped out to safeguard the common heritage. □

Paul G. Bahn is a freelance writer in archaeology at 428 Anlaby Road, Hull HU3 6QP, UK.

Evolutionary trees

What was the first living cell?

David Penny

WHAT can we learn about the nature of the first organisms from the DNA sequences of living organisms? The paper by Lake on page 184 of this issue¹ attempts to address this question, and concludes that the last common ancestor (shared by all living organisms) was similar to an 'eocyte' — sulphur-metabolizing organisms that are thermophilic (living in a hot environment). Such a conclusion challenges the favourite view of the past 50 years that the first organisms were heterotrophs living in a 'biochemical soup' that satisfied both their energy and substrate requirements from prebiotic organic molecules.

With such a bold claim, and considering the timescale of several billion years, an analysis of the methodology used by Lake is crucial. Now that DNA sequencing is becoming routine, reconstruction and testing of evolutionary trees is the main remaining challenge. Lake's conclusion relies on the now familiar assumption² that ribosomal sequences, by being universal, highly conserved and readily sequenced in living organisms, are the best candidates for detecting early divergences.

The important issue now is the reliability of trees reconstructed from DNA sequences. It is known that most methods of reconstruction can, for various reasons, give an incorrect tree, even when extremely long sequences are used. The results are particularly prone to error³ when organisms have different rates of evolution (nucleotide substitution in the sequence). The problem is more general than this in that an incorrect tree can occur even with equal rates of evolution⁴. Lake's new 'evolutionary parsimony' method is designed to minimize this problem, by using the information that in many

sequences transitions are more frequent than transversions. The method is thereby less affected by variations in the rate of evolution. It is one of several newer 'quartet' methods that consider taxa in groups of four^{4,5}. Evolutionary parsimony, theoretically and by simulation⁵, is more robust than many earlier methods, so it is necessary to consider Lake's new tree in detail.

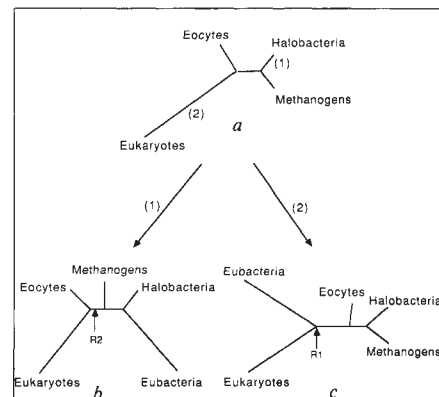
A formal analysis of the differences between the present tree¹ and earlier ones starts by considering the unrooted tree. If, for the moment, we omit the eubacteria, then there are four groups (eukaryotes and three archaeobacterial groups: eocytes, halobacteria and methanogens). Different methods appear to agree on these four groups and give the tree shown in *a* in the figure. Lake's result¹ has the fifth group (eubacteria) coming into the edge marked '1' to give *b* (the eocyte tree). Earlier methods joined the eubacteria at '2' to give *c* (the archaeobacterial tree).

When there are, for whatever reason, two longer edges such as for the eukaryotes and eubacteria, then most tree-building methods would give the tree in *c* in the figure, even if the tree in *b* was the correct one. This is because of the larger number of nucleotide changes on these two lineages. (Why there seems to be a faster rate in eubacteria and eukaryotes is an area that warrants investigation.) There is a strong tendency for 'long edges to attract' during tree building, and the difference in lengths is exaggerated if there have been different rates of molecular evolution. The problem is in distinguishing the cases when *c* is correct from the cases when it is an artefact and *b* is correct. Evolutionary parsimony is designed to help solve such cases.

Different methods of reconstruction

also result in differences in the placement of the root of the tree. Nearly all methods build an unrooted tree first, and then insert the root. (Cluster analysis methods can be considered to build an unrooted tree, and then to add the root into the last edge as it is formed.) There are five main methods of reconstruction.

The first evolutionary tree² placed the root at an internal node (R1 in the figure), leading to the eubacteria–archaeobacteria–eukaryote trichotomy. All subsequent work has supported the concept that the three archaeobacterial groups are distinct from both eukaryotes and other bacteria: it is the relationship between the three groups that is still controversial. The metabolic properties of the original cell (the progenote) are not specified in this method². The second, and perhaps most popular, view is that the eukaryotic nucleus



a, The unrooted tree for the four groups, excluding eubacteria. Lake¹ joins eubacteria to the edge marked (1) in *a* to give *b*. Earlier methods² joined eubacteria into edge (2) to give *c* and initially placed the root of the tree at a node (R1) in the tree. Lake's method places the root on the branch R2.

is derived from one of the archaeobacterial groups⁶. Lake's suggestion in this issue can be considered a specific version of this hypothesis. Third, Lake concludes that the root lies between the eocytes and the methanogens (R2 in the figure). His criterion is to minimize the number of changes to the molecular clock required to allow the data to meet the relative rate test. This method uses an objective criterion but is perhaps the most difficult to test. Fourth, Cavalier-Smith⁷ places the root within the eubacteria (partly to account for the double membrane and wall structure of these organisms) and suggests these earliest organisms were photoheterotrophs, using light as the energy source but still requiring organic molecules from the environment.

Finally, another alternative is possible if some introns are primitive⁸. This would account for observations such as the average exon size⁹ being similar to the Eigen limit for non-enzymatic replication of RNA¹⁰. The prokaryotes generally would be derived from a group that developed a

single replicon for DNA replication, giving a selective advantage for the elimination of introns by reducing the time for replication. Introns occur outside the eukaryotes but this is consistent with some introns being ancient if the non-eukaryote introns have a structural and regulatory role¹¹, and could not be eliminated without affecting survival. This hypothesis would place the root between eukaryotes and the other groups. It would still be consistent with Lake's conclusion of the last common ancestor being a sulphur-metabolizing thermophile, but provides less strong support for this view.

The conclusion from all these studies is that there is potential in reconstructing trees for all the main groups of living organisms. It is perhaps the best approach we have at present for learning the energy source of the last common ancestor. Lake's paper does help to set a higher standard for tree building from DNA sequence data. It is no longer acceptable to throw sequences through any available tree-building method and to publish the results. Some analysis of the reliability of the methods is essential, whether it be by statistical methods such as repetitive sampling for convergence, jack-knifing or bootstrapping^{12,13}, or an understanding of the mathematical limits of a particular tree-building method.

Newer methods have already been developed¹⁴ that allow more complex models of evolution, for example, allowing nucleotide positions to have different probabilities of change. There is a need for more powerful methods to handle other variations, such as changes in GC composition. The limits of methods that handle taxa in groups of four need to be evaluated because some have limitations that can be overcome by including additional sequences in each analysis⁴. More work is required on testing the reliability of tree-construction methods. The prize is not whose tree is correct, but a better understanding of the last common ancestor of all living organisms and, by implication, elucidation of the origin of life. □

1. Lake, J.A. *Nature* **331**, 184–186 (1988).
2. Woese, C.R. & Fox, G.E. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5088–5090 (1977).
3. Felsenstein, J. *Syst. Zool.* **27**, 401–410 (1978).
4. Penny, D. *et al. Cold Spring Harb. Symp. quant. Biol.* (in the press).
5. Li, W.H. *et al. Cold Spring Harb. Symp. quant. Biol.* (in the press).
6. van Valen, L.M. & Maiorana, V.C. *Nature* **287**, 248 (1980).
7. Cavalier-Smith, T. *Ann. N.Y. Acad. Sci.* **503**, 17–54 (1987).
8. Gilbert, W. *et al. Cell* **46**, 151–152 (1986).
9. Blake, C.C.F. *Nature* **306**, 535–537 (1983).
10. Eigen, M. & Winkler-Oswatitsch, R. *Naturwissenschaften* **68**, 282–292 (1981).
11. Shub, D.A. *Cold Spring Harb. Symp. quant. Biol.* (in the press).
12. Felsenstein, J. *Evolution* **39**, 783–791 (1985).
13. Penny, D. & Hendy, M.D. *Molec. Biol. Evol.* **3**, 403 (1986).
14. Olsen, G. *Cold Spring Harb. Symp. quant. Biol.* (in the press).

David Penny has just moved from the Department of Zoology, University of Cambridge, UK to the Department of Botany and Zoology, Massey University, Palmerston North, New Zealand.

Neurobiology

More light on brains

Larry Cohen

APPLICATIONS of optical methods for monitoring activity in the vertebrate brain *in vivo* have recently been reported in many laboratories^{1–12}. Some of the results, which depend on earlier developments^{13–15}, are unexpected and striking. In his report on page 166 of this issue¹², J.S. Kauer bridges the wide gap left by two earlier protocols. One of these uses an array of individual silicon photodiodes followed by parallel amplification, and has good time resolution (1 ms) but very low spatial resolution (10×10 pixels). The second protocol¹ uses a serial-readout video camera. Because of the need for averaging many frames, the temporal resolution is low (10,000 ms) but the spatial resolution is very high (128×120 pixels).

In his paper in this issue, Kauer presents results with good spatial resolution (128×128 pixels) and with relatively fast time resolution (30 ms, assuming no persistence). This allows him to examine, in some detail, both the spatial and the temporal spread of signals in the salamander olfactory bulb in response to olfactory nerve stimulation. When he compares the bulb response to stimulation of olfactory nerve fascicles from widely separated regions of the olfactory epithelium, he finds that the response areas and signal sizes are similar.

Four related questions arise from these results: (1) how has Kauer done so well; (2) what are the possibilities for improving the signal-to-noise ratio; (3) what are the optical, cellular and physiological origins of the long-lasting cortical signals found by Blasdel and Salama¹; and (4) will it be possible to determine the origins of the optical signals in three dimensions?

As Kauer points out, his good spatial and temporal resolution was achieved because the fractional fluorescence changes ($\Delta F/F$) in salamander olfactory bulb are the largest (1 per cent) thus far observed in vertebrate brain. The signal-to-noise ratio in a shot-noise limited measurement is proportional to the fractional intensity change and to the square root of the number of photons detected. Either increasing temporal resolution or increasing spatial resolution will reduce the number of detected photons and thereby reduce the signal-to-noise ratio. Thus there is a temporal–spatial tradeoff; but larger fractional changes will allow a larger temporal–spatial product. Because most brain activity will result in signals substantially smaller than 1 per cent, attention needs to be paid to reducing the noise and to increasing the signal.

Noise. Some noise sources (shot noise,

dark noise, quantization noise) are well understood. An array of individual detectors followed by parallel amplifiers is close to ideal, but the resolution (number of pixels) has been low. Efforts are under way in several laboratories to increase the number of detectors to as many as 2,500. On the other hand, camera systems with serial readouts are further from ideal. They have larger dark noise, quantization noise may be detectable, and shot noise will be relatively large because the number of pixels is large and the saturation level is low. But the commercial and military importance of these camera systems means that improvements are likely.

Other troublesome sources of noise are movement artefacts from the heartbeat and breathing. One possible approach to reducing these artefacts would be to subtract simultaneous measurements at two wavelengths, one where there is a dye signal and one at a longer wavelength which would detect only the movement^{16,17}. A second approach would be to use one of the recently developed mammalian preparations which allow a continuous (rather than pulsatile) perfusion.

Signal size. Surprisingly, little attention is given to attempts to find dyes with larger signals: there are only three laboratories in which new voltage-sensitive dyes are being synthesized. Another area where improvements in signal size could be made is in dye delivery. The dyes that work in the cortex are much less hydrophobic and have much smaller signals than those which give the best signals in squid axons. Methods for delivering these more hydrophobic dyes to cortical membranes might allow larger signals to be recorded.

Origins of the slow signals. Kauer presents his results as a series of 16 images spaced in time every 33 ms. He made them by subtracting a series that did not include a stimulus from one that did. In contrast, the method of Blasdel and Salama¹ provides only a single image from each experiment, and their simplest published pictures are the subtraction of two activity frames (using two different visual stimuli) rather than the subtraction of background from activity. For other results, such as orientation gradients, Blasdel and Salama used many frames to generate a single picture. The use of information from trials made at several orientations may explain why these authors could report a spatial resolution (tens of microns) that was much better than expected from light-scattering measurements (hundreds of microns)¹⁸.

The Blasdel–Salama results¹ appear reproducible both within Blasdel's labora-

tory, where similar signals have now been seen in eight monkeys, and in reports by Grinvald *et al.* (ref. 2 and D.Y. Ts'o, R.D. Frostig, E.E. Lieke & A. Grinvald, unpublished results), who have found apparently similar signals. Grinvald and collaborators, however, obtained their signals in the absence of a voltage-sensitive dye, which raises the possibility that some portion of the Blasdel–Salama signals do not arise from the dye but rather from changes in light scattering or intrinsic absorption. The cellular origins (neurons, glia or blood vessels) of the Blasdel–Salama signals are unknown, nor is it known whether these signals represent the summation of a fast signal that is coincident with neuron activity; or one of several substantially slower intrinsic signals^{2,17,19}.

Three-dimensional localization. The lack of three-dimensional localization is an important missing element in optical recordings from intact brain. The use of confocal microscopy might provide this kind of data.

Optical measurements already achieve temporal–spatial products that make them interesting to vertebrate neurobiologists and neurosurgeons. It remains to be seen whether the signal-to-noise ratios presently available are large enough to answer certain immediately obvious

questions. For example, will Kauer (or others) be able to detect odour-specific activity without further development of the techniques? Because the application of optical measurements to intact brain is less than five years old, many possibilities for improvement have yet to be explored. □

1. Blasdel, G.G. & Salama, G. *Nature* **321**, 579–585 (1986).
2. Grinvald, A., Lieke, E., Frostig, R.D., Gilbert, C.D. & Wiesel, T.N. *Nature* **324**, 361–364 (1986).
3. Cattarelli, M. & Cohen, L.B. *Soc. Neurosci. Abstr.* **12**, 1357 (1986).
4. Lektin, A. & Osmol, M. *Biophys. J.* **49**, 414a (1986).
5. Kauer, J.S., Senseman, D.M. & Cohen, L.B. *Brain Res.* **418**, 255–261 (1987).
6. Orbach, H.S. & Van Essen, D.C. *AVRO Abstr.* **28**, 197 (1987).
7. Evans, D. & Smith, J.C. *Brain Res.* **409**, 350–357 (1987).
8. Arieli, A. *et al. Abstr. Soc. Neurosci.* **13**, 11 (1987).
9. Senseman, D.M., Garcia, C.L. & Wayner, M.J. *Abstr. Soc. Neurosci.* **13**, 1411 (1987).
10. Tootell, R.B.H. & Blasdel, G.G. *Abstr. Soc. Neurosci.* **13**, 2 (1987).
11. Kiorpes, L. & Blasdel, G.G. *Abstr. Soc. Neurosci.* **13**, 1243 (1987).
12. Kauer, J.S. *Nature* **331**, 166–168 (1988).
13. Davila, H.V. *et al. Nature new Biol.* **241**, 159–160 (1973).
14. Salzberg, B.M., Grinvald, A., Cohen, L.B., Davila, H.V. & Ross, W.N. *J. Neurophysiol.* **40**, 1281–1291 (1977).
15. Orbach, H.S., Cohen, L.B. & Grinvald, A. *J. Neurosci.* **5**, 1886–1895 (1985).
16. Baylor, S.M., Chandler, W.K. & Marshall, M.W. *J. Physiol., Lond.* **331**, 105–137 (1982).
17. Jobbs, F.J., O'Connor, M., Vitale, A. & Vreman, H. *J. Neurophysiol.* **34**, 735–749 (1971).
18. Orbach, H.S. & Cohen, L.B. *J. Neurosci.* **3**, 2251 (1983).
19. Cohen, L.B. *Physiol. Rev.* **53**, 373–418 (1973).

Larry Cohen is in the Department of Physiology, Yale University School of Medicine, New Haven, Connecticut 06510–8026, USA.

Plate tectonics

Fate of subducted lithosphere

Peter Olson

THE global pattern of subduction zones, as revealed by deep seismicity, is the most direct evidence for convective transport in the mantle. It provides a remarkably unambiguous picture of the instantaneous positions of lithospheric slabs, sinking into the upper mantle under their own weight, to depths of about 650 km. Unfortunately, the Earth's seismicity ends there, so only the most recently subducted portions of slabs are directly visible. One of the most controversial issues in geophysics concerns the location of the older, aseismic portions of slabs, which must have been sinking into the mantle for as long as plate tectonics has operated. On page 131 of this issue¹, Ringwood and Irifune present a summary of density measurements on high-pressure mantle phase transformations that suggests some slab material may become buoyantly trapped in a layer between 600 and 700 km deep, and in doing so may have progressively stratified the mantle during Earth's history.

The driving forces for plate tectonics come primarily from the negative buoyancy of subducted oceanic lithosphere, which has cooled and contracted for an average of 80 million years, during which time it was at the surface.

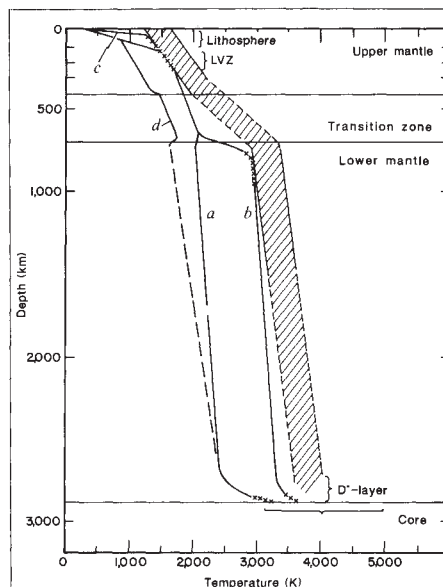
Upon subduction, oceanic lithosphere is 2–3 per cent more dense than the surrounding mantle. Torques imparted by sinking slabs establish a mantle-wide stress field that is largely responsible for plate tectonics. If the lithosphere and the rest of the mantle had precisely the same

composition, slabs would sink to the core–mantle boundary, become reheated and either become mixed back into the mantle, recirculate back to spreading centres, or both.

Although appealingly simple, this picture omits complications arising from differences of bulk composition within the mantle and within the lithosphere itself. Differences of bulk composition between the upper and lower mantle have long been advanced as the explanation for the seismicity terminus at about 650 km. In this model, slabs are unable to penetrate into the lower mantle, which is intrinsically more dense². There is no proof, however, that such a compositional difference exists. In fact, the preferred upper-mantle composition, a 2:1 mixture of olivine and pyroxene, elevated to the appropriate pressures, can account for 98 ± 2 per cent of lower-mantle elasticity and density³.

Instead of trying to resolve compositional differences between the upper and lower mantle, Ringwood and Irifune examine the role of the transition zone in altering the course of subduction. Between roughly 12 and 26 GPa, the pressure equivalents of depths of 400–750 km, mantle minerals such as olivine, pyroxene and garnet undergo a sequence of solid-state phase transformations, to a mixture of more dense silicate perovskite and oxide structures. The pressure at which the critical perovskite-forming reactions occur is increased by lowering the temperature or by presence of minor abundant cations such as calcium and aluminium. This is potentially important in subducted slabs, which are both colder and more highly differentiated than the surrounding mantle.

Slabs are composed of laminations, the two most significant being the 6–7 km layer of mid-ocean-ridge basalt (MORB) crust overlying a 30-km residuum layer



Possible mantle geotherms and sites of active chemical differentiation. Curves *a* and *b*, spherically averaged geotherms, appropriate for a homogeneous and a stratified mantle, respectively. Representative geotherms beneath oceanic spreading centres (*c*) and in subducted slabs (*d*) shown for comparison. The slab geotherm is projected to the core–mantle boundary, assuming that slabs may occasionally sink into the lower mantle. Horizontal lines mark 400-km and 650-km seismic discontinuities and the core–mantle boundary. Hatching, range of measured solidus temperatures for abundant upper- and lower-mantle minerals (solid borders) and their estimated range in the transition zone and deep in the lower mantle (dashed borders). Bracket, possible outermost-core temperatures. Active differentiation of slab material (crosses) can occur where geotherm approaches or exceeds the solidus, as in the seismic low velocity zone (LVZ), in the D''-layer above the core–mantle boundary, and perhaps¹ at the base of the transition zone, implying a geotherm similar to curve *b*.

(harzburgite) left over after basalt extraction. A cornerstone of mantle petrology is the assumption that the parent material of these two products represents the bulk composition of the upper mantle. Ringwood calls this material pyrolite; he and Irifune use it as the standard with which slab properties are compared. It is worth emphasizing that no field geologist has ever found an outcrop of pyrolite; it is a hypothetical rock type with a composition chosen to account for MORB genesis by partial melting beneath spreading centres.

Ringwood and Irifune find that within the interval between 50 and 650 km, MORB is on average 3.5 per cent more dense, and harzburgite 1.5 per cent less dense, than pyrolite at the same temperature. These differences may not be relevant to slab behaviour in the upper mantle because the net density anomaly of the laminate is nearly zero and is swamped by the density excess resulting from thermal contraction.

In addition, extremely high viscosity in the cold slab prevents the layers from separating, but near the base of the transition zone the density relationships change. The higher aluminium and calcium content in MORB prevents perovskite formation for 1–2 GPa beyond the pressure at which it forms in pyrolite, and within that depth interval (near 700 km) former oceanic crust becomes less dense than surrounding mantle at the same temperature. Conversely, depletion of aluminium and calcium in harzburgite makes it locally more dense than pyrolite in the same region. At still higher pressure, the density relations found in the upper mantle prevail again. Thus, buoyancy conditions are right for gravitational segregation of the lithosphere in a channel between the upper and lower mantle.

Is this separation likely to occur? Is there independent evidence that it does? Separation of adjacent strips of material is a mechanical problem. The average subsolidus viscosity of the mantle is near 10^{21} Pa s and even higher within cold slabs, providing strong adhesion between adjacent laminations, so that under normal conditions separation is extremely slow. The situation is made worse by streamline mixing associated with convective flow, which acts to stir the lithosphere into the mantle, probably on a timescale of a few hundred million years⁴. To reduce adhesion and to allow separation to occur, slab temperatures must be raised to near the fusion point where the viscosity becomes low enough to permit differential motions on a small scale. Only in a few regions of the mantle is this likely: the low-velocity zone (LVZ), the D"-layer just above the core boundary, and perhaps at the base of the transition zone (see figure).

The LVZ, a region of partial melts, is the site of the original differentiation producing MORB. At the base of the

mantle, the geotherm curves sharply through the D"-layer to meet (or exceed) the outer core liquidus. This layer has already been proposed as a site for segregation of slabs subducted into the lower mantle⁵. For Ringwood and Irifune's proposal to work, the geotherm must lie close to the solidus near 700 km, similar to curve *b* in the figure, implying the existence of a diffusive thermal gradient in the anomalous layer, and elevated temperatures in the lower mantle. The evidence for geotherm *b* is weak. There is nothing to indicate the existence of a low-viscosity channel in that region. In fact, there is good evidence that the viscosity increases from upper to lower mantle⁶, suggesting, if anything, that the geotherm diverges from the solidus, as indicated by curve *a* in the figure.

This is not to imply there is not independent support for transition-zone differentiation. As Ringwood and Irifune point out, the presence of thin lenses of former ocean crust could explain the

anomalous seismic reflectivity of the transition zone, as revealed by compressional-wave reflections from the underside of the 650-km discontinuity. It is fair to say there is evidence in favour of and against active differentiation in the transition zone, but neither is conclusive. On the other hand, such regions should be recognizable by virtue of low-viscosity, high seismic attenuation and lateral heterogeneity and other observable characteristics, so Ringwood and Irifune's proposal is certainly testable using available geophysical methods. □

1. Ringwood, A.E. & Irifune, T. *Nature* **331**, 131–136 (1988).
2. Lees, A.B., Bukowski, M. & Jeanloz, R. *J. geophys. Res.* **88**, 8145–8159 (1983).
3. Knittle, E., Jeanloz, R. & Smith, G. *Nature* **319**, 214–216 (1986).
4. Allegre, C.J. & Turcotte, D.L. *Nature* **323**, 123–127 (1986).
5. Hofmann, A.W. & White, W.M. *Earth planet. Sci. Lett.* **57**, 421–436 (1982).
6. Hager, B.H. *J. geophys. Res.* **89**, 6003–6015 (1984).

Peter Olson is in the Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland 21228, USA.

Molecular neurobiology

A Rosetta stone for K channels

William S. Agnew

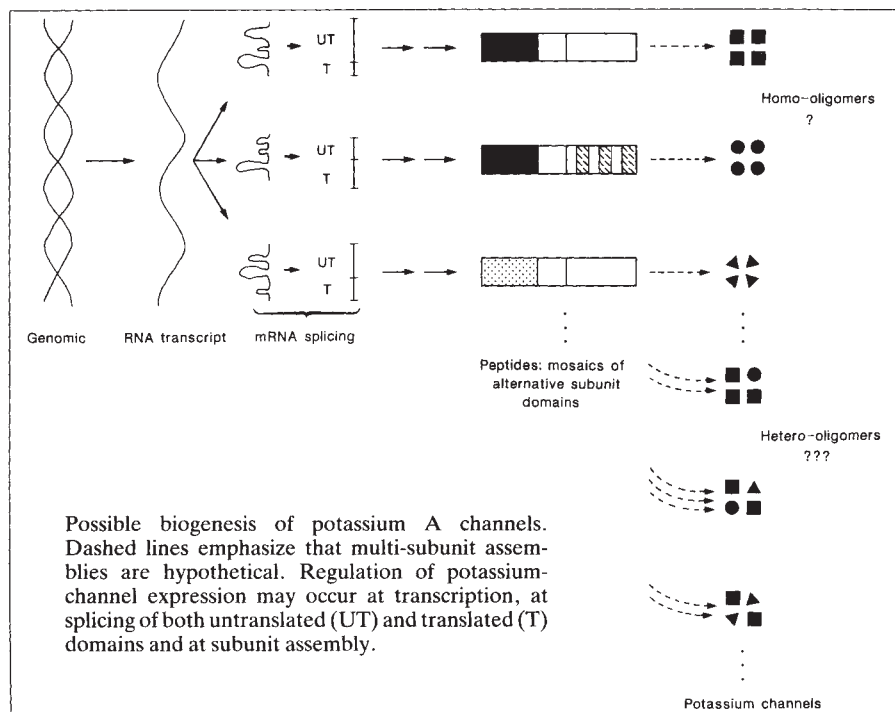
A NERVOUS system lacking ion channels would be a very quiet place indeed. It would seem to follow that by focusing on what might be a comparatively small number of channels and enzymes that mediate and modulate electrical excitability, we might gain simplifying insights into a very complex subject. Data provided by patch-clamp electrodes and an arsenal of toxins, drugs, transmitters and second messengers fuel our suspicion that the diversity of channels, although rich, is probably comprehensible.

However, as we begin to explore the genetic library for the precise identities and origins of these channels, we should anticipate surprises. The picture emerging from an examination^{1,2} of potassium A channels in the fruitfly *Drosophila*, including the companion papers from the Jan laboratory on pages 137 and 143 of this issue^{3,4}, is of a marvellously flexible set of molecular building blocks and assembly instructions for the construction of excitable membranes.

The X-linked *Shaker* mutations were among the earliest behavioural variants identified in *Drosophila*. About 10 years ago it became apparent that the characteristic spontaneous excitability, poor motor control and abnormal synaptic efficiency result from defects in transient, voltage-activated potassium A channels encoded by this locus^{5–7}. By controlling repolarizing outward currents, these channels markedly affect action-potential threshold, frequency and duration. The currents in

both muscle and neurons are affected. Subsequently, using chromosome walking and hybrid dysgenesis, the approximately 120 kilobases of genomic DNA encompassing the *Shaker* locus were cloned^{1,8,9}. Now, hybridizing complementary (c)DNAs prepared from messenger (m)RNA can be screened and cloned, and the genomic sequences contributing to the mRNA transcripts identified (see figure).

The first complete sequence for the protein-coding region of a *Shaker* clone was recently reported², providing direct comparisons with other ion-channel peptides. Previously, voltage-sensitive sodium-channel and dihydropyridine-binding (calcium-binding) proteins had been found to be large peptides containing four separated, internally homologous domains encompassing all the potentially membrane-spanning hydrophobic sequences^{10–13}. This suggested an organization of $M_r \sim 30,000$ 'pseudo-subunits' linked in a chain, probably forming a tetrameric rosette in the membrane bilayer. A stereotypic pattern in each pseudo-subunit of six fairly hydrophobic segments included one decorated at every third residue with an arginine or lysine. These 'S-4' segments almost certainly contribute to the voltage-sensing mechanisms. The *Shaker* protein showed striking homology to these peptides, but was much smaller, corresponding to only one pseudo-subunit linked to hydrophilic carboxy- and amino-terminal sequences. This suggests that, if the *Shaker* protein is a component of the



potassium channel, the channel might be an assembly of separate subunits, each containing membrane-spanning regions similar to those of other voltage-sensing proteins. Thus, it is essential to determine whether the peptide is truly a component of the potassium channel and whether other subunits are involved. Also, if voltage-sensing channels share a common evolutionary ancestor, why are the subunits of sodium channels and dihydropyridine receptors linked and those of the *Shaker* channel separate? Answers to these questions are now emerging.

The Jan group has now isolated and sequenced seven *Shaker* cDNAs, together with the corresponding genomic sequences. They discover five different species, four of which differ in the open reading frame that encodes the proteins. It is strikingly clear that the five different transcripts are derived by alternative mRNA splicing.

In one of the papers in this issue⁴ by this group, Timpe *et al.* transcribed two of the *Shaker* cDNAs *in vitro* and expressed them separately in *Xenopus* (frog) oocytes, yielding characteristic voltage-activated, rapidly inactivating, 4-aminopyridine-sensitive potassium currents. These currents show clear, if subtle, differences. Thus, if the *Shaker* channels are multimeric, homo-oligomers of at least two and possibly all four peptides form functional channels. In Northern blot analyses of poly(A⁺) mRNA, Schwartz *et al.* in the other paper in this issue³ demonstrate that more than five mRNA species are synthesized. A guess of ten would not be out of order, and thus potentially ten distinct channels may derive from *Shaker*. But the situation could be distinctly more complex if, by loose analogy with the sodium-

channel and dihydropyridine-receptor proteins, hetero-oligomers are functional and occur normally. If the *Shaker* channels are tetrameric assemblies of several — if not all possible — subunit combinations, the upper limit of channel subtypes could extend from dozens to many thousands. Even if this unrestricted reassortment does not naturally occur, hetero-oligomers expressed artificially could help unravel the association of specific domains with the mechanisms of potassium-channel function.

The primary structure of the *Shaker* proteins described in this issue^{3,4} derives from the splicing of 11 DNA exons in a mix-and-match pattern. The previously reported ShA1 cDNA predicts a smallish protein, of *M_r* 70,200, containing the six potential membrane-spanning hydrophobic sequences, including the S-4 sequence, which is the apparent signature of voltage-sensing proteins². There are extended hydrophilic sequences at the amino and carboxy termini. Comparing all four *Shaker* clones, it is apparent that nearly all the hydrophobic segments derive from eight closely linked exons which are identical for all four proteins. The variable carboxy termini are further derived from either of two sets of two exons, and the amino termini from one of three exons. What is intriguing is that the amino and carboxy elements appear to vary independently.

Thus, ShA1 and ShC1 have identical carboxy but different amino termini, whereas ShB1 and ShB'1 have the same amino terminus as ShA1 but different carboxy sequences. ShD1 has the ShB1/ShB'1 carboxy sequences but an amino terminus different from them and from ShA1 and ShC1. Notably, the carboxy

termini are sufficiently similar that, together with the conserved membrane-spanning regions of the proteins, hetero-oligomeric associations seem at least possible. Could this apparently independent shuffling of domains correspond to recombination of distinctive properties associated with voltage-sensitivity, kinetics, second-messenger regulation and topographical distribution? How many alternatives remain?

There is yet more to this story. The independent segregation of protein-coding exons is accompanied by variations in the untranslated regions. ShB1 and ShB'1 specify identical peptides but markedly different 5' untranslated regions, suggesting mechanisms for regulation of translation. Jan and co-workers note that their cDNAs are not full length and the mRNAs found by Northern blots correspond to species from 5.5 up to 9.5 kilobases, all explicitly derived from the *Shaker* locus. Given that the structural sequences encompass less than 2 kilobases, what are the other 60–80 per cent of the nucleotide sequences doing?

Clearly, the *Shaker* locus provides for maximum flexibility in potassium-channel expression. The commitment to form a channel subtype is delayed at least to the splicing stage, and possibly until translation or even subunit assembly. Obviously, the commitment steps must be regulated. The splicing mechanism that selects which combination of exons contributes to the coding and untranslated sequences very possibly responds to developmental and neuromodulatory regulation.

These findings with the *Shaker* locus will no doubt now be extended to the vertebrate nervous system. Voltage-sensitive sodium channels have been strongly conserved through the 600 million years of evolution separating mammals and fruit-flies¹². Potassium A channels occur in the mammalian nervous system and can be expressed in frog oocytes, and it seems certain the *Shaker* clones will facilitate their isolation. This may begin the deciphering of an even more intriguing chapter in the story of potassium channels. □

1. Papazian, D.M., Schwarz, T.L., Tempel, B.L., Jan Y.N. & Jan, L.Y. *Science* **237**, 749–757 (1987).
2. Tempel, B.L. *et al.* *Science* **237**, 770–775 (1987).
3. Schwartz, T.L. *et al.* *Nature* **331**, 137–142 (1988).
4. Timpe, L.C. *et al.* *Nature* **331**, 143–145 (1988).
5. Jan, L.Y. & Jan, Y.N. *J. Physiol., Lond.* **262**, 189 (1976).
6. Tanouye, M.A., Ferrus, A. & Fujita, S.C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6548–6552 (1981).
7. Salkoff, L. & Wyman, R. *Nature* **293**, 228–230 (1981).
8. Kam, A., Iverson, L.E. & Tanouye, M.A. *Cell* **50**, 405–413 (1987).
9. Banmann, A., Krah-Jengens, F., Muller, R., Muller-Holtkamp, A. & Pings, O. *EMBO J.* **6**, 3419–3429 (1987).
10. Noda, M. *et al.* *Nature* **312**, 121–127 (1984).
11. Tanabe, T. *et al.* *Nature* **328**, 313–318 (1987).
12. Salkoff, L. *et al.* *Science* **237**, 744–748 (1987).
13. Agnew, W.S. *Nature* **328**, 297 (1987).

William S. Agnew is in the Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, Connecticut 06510–8026, USA.

Chemical physics

How small is a solid?

Tony Stace

How large does a collection of atoms or molecules have to be before it begins to adopt the properties and features of a solid? Experiments on gas phase clusters consisting of small numbers of atoms or molecules (10–1,000) are providing estimates of the minimum size for a solid. The most recent study, by K. Rademann *et al.*¹, concentrates on measuring the ionization potentials of mercury clusters containing up to 70 atoms. These results show that there is a gradual transition from essentially atomic behaviour for small clusters (5–10 atoms), to metallic behaviour for clusters with 60–70 atoms.

A crystalline solid has several properties that can clearly be identified as characteristic of the bulk material rather than of its constituent atoms or molecules. Examples are the phonon spectrum, a work function, and an electronic band structure. At the same time, however, it is possible to see that such properties have emerged from a unification of individual constituents, and that many are a direct extrapolation of an equivalent atomic or molecular property. Thus, the solid work function is like the ionization potential of a molecule, the metal conduction band is like an atomic orbital, and the solid heat of vaporization is like the binding energy between a pair of atoms or molecules.

Single mercury atoms have a closed-shell (s^2) configuration of electrons, and if the relative energies of the occupied outer s -orbital and the first unoccupied p -orbital were not influenced by clustering, then bulk mercury would most probably be an insulator. In practice² (see Fig. 1), the s - and p -orbitals spread and move towards one another as the cluster size increases, and it appears that for 70 atoms there is sufficient overlap for a conduction band to form. It is normal practice, when comparing the ionization potentials of clusters with the bulk work function, to make a Coulomb energy correction which allows

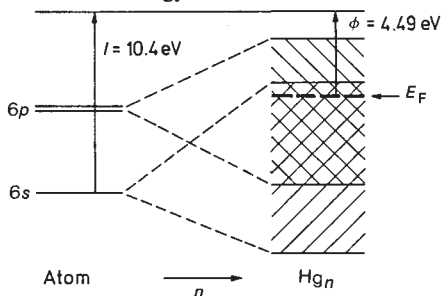


Fig. 1 Closing and spreading of the energy gap between the atomic 6s and 6p orbitals of mercury for increasing cluster size, n . I , atomic ionization potential; ϕ , work function and E_F , Fermi energy of the 'solid'. (From ref. 2.)

for the fact that in a cluster the remaining positive charge is confined to a finite rather than an infinite volume.

With such a correction, the measured ionization potentials of Rademann *et al.* compare very favourably with the work function for solid mercury. An ionization potential, however, is the energy difference between the electronic ground state of a neutral and the corresponding ion. Therefore, the magnitude of the energy difference could be influenced by some property of an ion cluster which is not necessarily associated with the formation

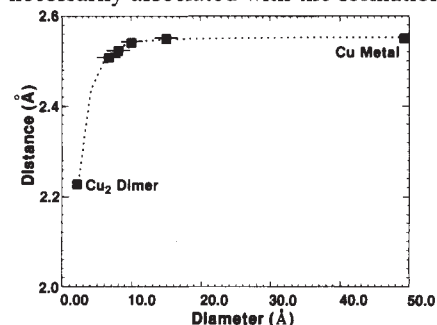


Fig. 2 Variation in internuclear distance for copper clusters as a function of cluster size. (From ref. 7.)

of band structure. One such property might be the cluster's ability to form a trapped hole (electron vacancy); here the enhanced stability gained from trapping can lead to a considerable lowering of the ionization potential. Argon ion clusters are a good example of such behaviour³.

Brechignac *et al.* (ref. 4 and personal communication) have studied the electronic structure of mercury clusters in a slightly different way from that discussed above. They used high-energy synchrotron radiation to promote an inner-shell, or core, electron to a valence state from which it induces autoionization. They argue⁴ that, although the core electrons remain unaffected by the development of bulk-like behaviour, the variation in photon energy necessary to promote electrons provides a measure of the changes in the valence states as a function of cluster size. Such experiments appear to be less sensitive to any special electronic characteristics of the resultant ions.

There is a slight difference between the two experiments in terms of the end result. Brechignac *et al.*⁴ suggest that 20 mercury atoms in a cluster are sufficient to initiate electronic band formation. In both experiments the mercury ion clusters undergo fragmentation between the time they are formed and when they are finally detected. Hence, the metal transition size

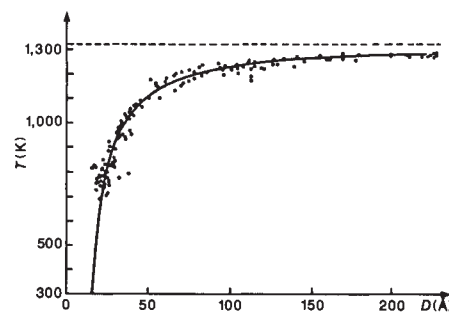


Fig. 3 Variation in the melting point T_m of gold clusters as a function of cluster size. Broken line, bulk melting point. (From ref. 8.)

could be slightly larger than is suggested by either group^{1,4}. Alternative experiments, designed to study the structures of clusters, also provide indicators of the onset of bulk behaviour. Solid argon has a face-centred-cubic (f.c.c.) crystalline structure (with octahedral symmetry), and yet both calculations⁵ and electron-diffraction experiments⁶ show that small (20–50) argon clusters have icosahedral (five-fold) symmetry. It is only if the clusters contain 100 or more argon atoms that the experiments show evidence of emergent octahedral symmetry.

In contrast, small copper clusters⁷ appear to adopt the f.c.c. symmetry of the bulk metal from Cu_{13} onwards. Copper clusters do deviate from bulk behaviour in the nearest-neighbour distance between atoms⁷. Figure 2 shows the variation in atom–atom distance as a function of cluster size. An internuclear distance appropriate for bulk copper is achieved only when the clusters contain 50–100 atoms.

It is not only the electronic and structural properties of clusters that exhibit size effects; even a concept as classical (and supposedly as constant) as the melting point of a crystalline solid is also subject to variation in clusters (strictly speaking we should not talk of melting in isolation from structure; both are intimately joined by intermolecular forces). Buffat and Borel⁸ have measured the melting points of small gold particles down to 20 Å in size (Fig. 3). Compared with the cluster sizes discussed above, it can be seen that melting requires 1,000 or more atoms per particle (taking 1.34 Å as the covalent radius for gold) even to approach the bulk value.

The considerable difference between this result and those discussed above can be linked to the ambiguity associated with defining a temperature for an isolated system (see my recent News and Views article⁹). Melting-point experiments, however, are usually performed on metal particles that are suspended on a substrate, and as a result are (we must presume) in thermal equilibrium with their surroundings. Alternatively, the end result could be influenced by the periodicity of the property being measured.

If it were necessary to put a value to the question "how small is a solid?", then

approximately 100 atoms would appear to be sufficient to form an adequate resemblance of bulk material. Such a figure, however, is arrived at from a consideration of what is essentially a very small subset of all the possible techniques for studying the properties of solids. It could be that the answer is sensitive to both the experimental method used and the property under analysis. For example, band structure may become established at quite an early stage in the growth of a particle, whereas the development of a phonon spectrum could require many thousands of atoms. Compared with solids and liquids, the study of clusters is still very much in its infancy. But over the next five

years we can expect to see significant advances in our understanding of how the properties of solids develop from their constituents. □

1. Rademann, K., Kaiser, B., Even, U. & Hensel, F. *Phys. Rev. Lett.* **59**, 2319–2321 (1987).
2. Benneemann, K.H. & Pastor, G. in *Microclusters* (eds Sugano, S. *et al.*) 54–62 (Springer, Berlin, 1987).
3. Dehmer, P.M. & Pratt, S.T. *J. chem. Phys.* **76**, 843–853 (1982).
4. Brechignac, C. & Cahuzac, Ph. *Z. Phys.* **D3**, 121–129 (1986).
5. Hoare, M.R. *Adv. chem. Phys.* **40**, 49–135 (1979).
6. Farges, J. *et al.* *J. chem. Phys.* **84**, 3941–3501 (1986).
7. Montano, P.A. *et al.* *Phys. Rev. Lett.* **56**, 2076–2079 (1986).
8. Buffat, Ph. & Borel, J.-P. *Phys. Rev.* **A13**, 2287–2298 (1976).
9. Stace, T. *Nature* **327**, 186–187 (1987).

Tony Stace is at the School of Chemistry and Molecular Sciences, Sussex University, Falmer, Brighton BN1 9QT, UK.

Human vision

Revealing the artist's touch

A.W. Snyder and H.B. Barlow

Is there a unifying conceptual framework for understanding human visual-information processing, one that explains the occurrence of illusions such as those of Ramachandran on page 163 of this issue¹? Or is it better to regard our visual system as a mosaic of unrelated strategies, one for every special task²? An animal needs to make rapid decisions about what objects are where in its surroundings, but it has only a few neurons at its disposal: this strongly suggests that the brain should first simplify processing for detection and recognition of objects, and then report its findings with the utmost economy. We propose that these two principles provide a unifying interpretation of existing perceptual data and will enable us to place new results in a proper perspective.

Simplification of processing

Scenes presented to the eye are not an unlimited succession of surprises, but conform to certain patterns. Visual processing would become more rapid and efficient if the brain had expectations about what is to be seen, and it would be plausible for these expectations to become incorporated into perceptual mechanisms during the course of evolution. The structure and sequence of visual processing would then be determined by these expectations, which should reveal themselves as illusions in contrived or unnatural situations. Now natural scenes, as distinguished from the infinity of contrived scenes, are composed of discrete objects illuminated principally from one direction: it is such attributes of objects as they occur in the natural world, that should, if we are right, be structured into visual processing. Most of these attributes are in fact already well known to artists because they have had to master them to create the illusion of the three-

dimensional world on a two-dimensional surface (see figure); relevant examples concern luminance gradients and colour.

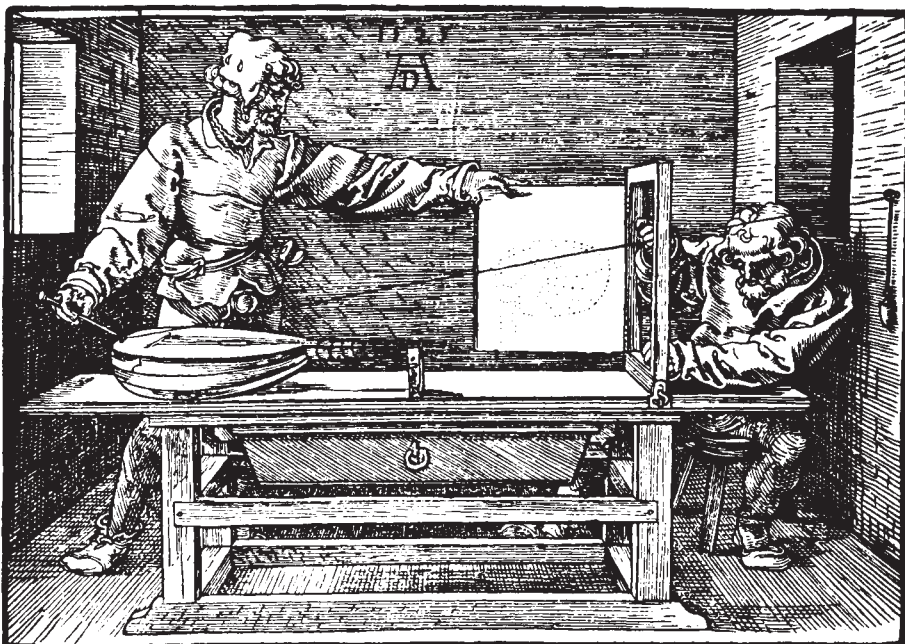
Objects are often visible in our natural world by contrasting with the background, so that luminance borders are a prime attribute used to demark their outline. Provided the brain has been given definite clues that there is an object (for instance by the clearly visible curved corners in Fig. 4c of Ramachandran's paper, on page 165), the existence of a border all round the object will be assumed, and one will be seen where no border actually exists. A reasonable generalization (that objects have borders) leads to an illusion in this artificially contrived situation, but the generalization would be highly advan-

tageous for rapid object detection in natural scenes. Various other illusions, such as the well-known Craik–Cornsweet illusion³, show that borders alone can establish the contrast of objects. Finally, the idea⁴ that the brain delays processing of internal regions until the border of the region is first established is supported by Ramachandran's Fig. 4c.

Consider next the artist's technique of shape from shading. This encapsulates the fact of our world that luminance gradients are a cue to the solid, three-dimensional percept. Indeed, Ramachandran shows¹ how the brain presupposes naturally occurring shading, with light from one direction. This assumption would lead to errors in a hypothetical world of diffuse or multiple light sources, but it simplifies the processing of natural scenes.

We could provide many other examples, such as perspective from texture gradients, but perhaps the most convincing economy of processing is demonstrated by colour vision. Isoluminance is rare in our natural world, partly because shading is an intrinsic property of three-dimensional objects. Thus, colour vision may have evolved for gaining attention and rapidly labelling objects that would be visible to an achromatic eye by virtue of their luminance borders. The brain could then simply associate coloured objects with their luminance borders, but some rather bizarre illusions are anticipated in unnatural situations. Stationary coloured objects can be seen to follow their associated (moving) luminant borders⁵, even when the associated border is illusory⁶.

Recognition of faces, and the perception of three-dimensionality, is extremely difficult in contrived situations where natural luminance gradients are replaced



Albrecht Dürer (1525). The use of object attributes to portray objects had to be learned by trial and error, and then taught to aspiring draughtsmen by the use of tricks and drawing aids.

by colour gradients, even though the information content is similar. Furthermore, colour contributes little to motion^{7,8} or to stereopsis^{8,9}. Clearly, human colour processing is greatly simplified by assumptions about the natural world, but at the cost of erroneous and inadequate appearances in artificially contrived situations.

Economy of reporting

The brain might thus accelerate its formulation of the percept by incorporating expectations about object attributes into the structure and sequence of processing. The report of this percept might also be hastened by economy of reporting. It is the objects themselves that are important to an animal, not the object attributes processed by the brain to formulate the percept. Artists know that well-executed object attributes are not apparent even after careful inspection; the viewer sees the object itself instead. The more familiar of these attributes include shape from shading, perspective from gradients of texture, and size invariance with distance, but painters through the ages had to learn their importance using trial and error¹⁰. Helmholtz¹¹ saw that "we are not in the habit of observing our sensations accurately, except as they are useful in enabling us to recognize external objects. On the contrary, we are wont to disregard all those parts of sensation that are of no importance so far as external objects are concerned". Thus, the brain reports only the object, suppressing information about the attributes used to find it. The reason, we suggest, is for economy of representation and reporting, an idea that may also be traced¹² back to the last century^{13,14}.

In summary, we think the brain has adopted two elementary simplifying principles, one for rapid processing of information and the other for rapid reporting. Many illusions are consistent with this interpretation, including those reported in this issue¹ by Ramachandran. It might henceforth be useful to view perception within this unifying perspective. □

1. Ramachandran, V.S. *Nature* **331**, 163–166 (1988).
2. Ramachandran, V.S. *Perception* **14**, 97–103 (1985).
3. Cornsweet, T.N. *Visual Perception* (Academic, London, 1970).
4. Osorio, D. *et al.* *Spatial Vision* **2**, 191–198 (1987).
5. Kolers, P.A. & von Granau, M. *Vision Res.* **16**, 329 (1976).
6. Ramachandran, V.S. *Nature* **328**, 645–649 (1987).
7. Ramachandran, V.S. & Gregory, R.L. *Nature* **275**, 55–56 (1978).
8. Lu, C. & Fender, D.H. *Invest. Ophthalm.* **11**, 482–490 (1972).
9. Livingstone, M.S. & Hubel, D.H. *J. Neurosci.* **7**, 3416–3468 (1987).
10. Gombrich, E.H. *Art and Illusion* (Phaidon, Oxford, 1960).
11. Helmholtz, H. *Handbuch der Physiologischen Optik* Vol. III (1867) (English tr. ed. Southall, J. Opt. Soc. Am., 1924).
12. Barlow, H.B. in *Sensory Communication* (ed. Rosenblith, W.) Ch. 13 (MIT, Cambridge, Massachusetts, 1961).
13. Mach, E. *The Analysis of Sensations*, translation of 1st German edn, 1886, by C.M. Williams (Open Court, Chicago, 1914).
14. Pearson, K. *The Grammar of Science* (Scott, London, 1892).

A.W. Snyder is Head of the Optical Sciences Centre at the Australian National University, Canberra 2601, Australia; H.B. Barlow is at the Physiological Laboratory, Cambridge CB2 3EG, UK.

Amorphous semiconductors

Making multilayered samples

Richard Friend

AMORPHOUS semiconducting materials are becoming increasingly popular for electronic applications, particularly for large-area devices such as solar cells. Multilayered structures, equally, are of great interest because of their applications in devices such as quantum-well lasers. The two technologies, however, have been hard to wed. On page 153 of this issue¹, M. Kawasaki *et al.* report a significant advance in this direction.

Multilayer semiconductor structures are usually fabricated as crystalline semiconductors such as GaAs/Ga_{1-x}Al_xAs, and are grown by techniques such as molecular beam epitaxy². Materials are now routinely grown with layer thicknesses down to 20 Å, without significant levels of interfacial defects. Properties investigated in these materials include the formation of quantum wells from the spatial confinement of carriers within the layers, and the spatial separation of charge between the different layers which can allow very high in-plane charge-carrier mobilities, and the easy observation of the quantum Hall effect.

The scope for building similar multilayer structures in amorphous semiconductors is much more limited; the defect structure at the interface is inherently difficult to control, and carrier mobilities are inherently low and unlikely to be improved by charge separation between layers. Most important, the preparation of amorphous (*a*-) silicon-hydrogen by glow discharge decomposition of silane is not easily adapted to the formation of multilayers such as *a*-silicon/*a*-silicon nitride. Kawasaki *et al.* have devised¹ a method of depositing *a*-silicon/*a*-silicon carbide multilayers with continuous plasma decomposition of silane (SiH₄), and a pulsed photochemical deposition of carbon tetrafluoride. The advantage of this method is that the superlattices can be deposited continuously, without the need to pump out the preparation chamber in between the deposition of successive layers.

Compositional superlattice semiconductors derive their novel electronic properties from the difference between the energy band gaps of the two types of layer; Ga_{1-x}Al_xAs, for example, has a larger gap than GaAs. Electrons at the bottom of the conduction band and holes (electron vacancies) at the top of the valence band of the GaAs layers are spatially confined parallel to the layer by the Ga_{1-x}Al_xAs layers on either side, so that although they behave as free particles within the plane, their wave functions perpendicular to it are described by the standard 'particle in a box' solutions of the

Schrödinger equation, with a series of discrete energy levels. These are seen in various optical experiments. Charge separation between layers can be achieved by including the dopant (that gives the substance *n*- or *p*-type character) in the larger-band-gap layers. When the dopants are ionized, the charge carriers liberated migrate to the smaller-band-gap layer. Free from the Coulomb potential of the ionized dopant, these carriers can exhibit huge mobilities, in excess of 10⁶ cm² V⁻¹ s⁻¹.

The properties of amorphous semiconductors are determined by the structural disorder: the random potentials which this sets up have the effect of localizing the electronic states near the band edges. Electronic transport then takes place through the hopping of charges between these spatially localized states, and electronic mobilities are low. In many amorphous semiconductors, the non-crystalline structure allows the local chemical bonding to adjust to local environment, around for example a dopant atom, so that substitutional doping is ineffective. Nevertheless, amorphous hydrogenated silicon formed by glow discharge from silane has a rather low density of states in the band gap, and can be doped in the conventional ways³. Various groups have fabricated compositional superlattices of this with amorphous silicon nitride, silicon carbide, silicon dioxide (all with larger band gaps than the amorphous silicon)^{4,5} and germanium⁶. These films have generally been grown by successive decomposition of silane (for the silicon layer) and a suitable gas mixture for the other layer type (silane and ammonia for silicon nitride).

Experimental evidence for formation of quantum wells in these structures is gradually accumulating. It is well known⁷ that the band gap of the silicon layer increases from its bulk value of 1.8 eV as the thickness of the layer drops below 50 Å. This is consistent with quantum-well formation, and quantitative agreement is achieved with reasonable values for the electron hole effective masses⁴, but the possibility that the 'interfacial' region extends throughout the 'silicon' layer for these very thin silicon layers is hard to exclude⁵. There are more specific experiments that do show evidence for formation of the quantum-well; these include optical measurements such as luminescence, and resonant-tunnelling conduction experiments which show the expected series of resonances into the quantized levels of the quantum wells⁴.

For the crystalline multilayers, much is

known about the bulk materials, and many of the properties of the superlattices were very accurately predicted using the parameters of the bulk materials before the superlattices were made and measured. Amorphous semiconductors have always been much more difficult to characterize and, in contrast, the recent measurements⁷ on multilayer structures have provided basic information about the bulk materials (effective masses in the conduction and valence bands). Improvements in preparation methods will doubtless lead to further advances in our basic understanding of these materials; an obvious extension of

the work¹ of Kawasaki *et al.* is to control the process of doping to fabricate doping superlattices⁷.

1. Kawasaki, M. *et al.* *Nature* **331**, 153–155 (1988).
2. Kelly, M.J. & Nicholas, R.J. *Rep. Prog. Phys.* **48**, 1699–1741 (1985).
3. Mott, N.F. & Davis, E.A. *Electronic Processes in Non-Crystalline Materials* (Oxford University Press, 1979).
4. Hirose, M. & Miyazaki, S. *Proc. Mater. Sci. Soc.* **3**, 502–509 (1987).
5. Kalem, S. *et al.* *Proc. 18th Int. Semiconductors Conf.* 747–750 (1986).
6. Tiedje, T., Wronski, C.R., Persans, P., & Abeles, B. *J. non-cryst. Solids* **77 & 78**, 1031–1040 (1985).
7. Döhler, G.H. *J. non-cryst. Solids* **77 & 78**, 1041 (1985).

Richard Friend is at the Cavendish Laboratory, Cambridge University, Madingley Road, Cambridge CB3 0HE, UK.

Inositol phosphates

Ins and outs of cell signalling

Jennifer Altman

It is only four years since inositol 1,4,5-trisphosphate (InsP_3) was established as a second messenger involved in the mobilization of calcium in cells¹. No doubt because calcium is a ubiquitous signal, fundamental to the control of processes ranging from regulating membrane potentials to cell proliferation and the development of organisms, interest in the inositol phosphates has burgeoned. The over-capacity audience at a recent meeting* and a lively subsequent discussion group debated how InsP_3 regulates calcium and the role of the more recently discovered inositol tetrakisphosphate (InsP_4 , ref. 2). But a new departure was also suggested — there are indications^{3,4}, one of which was recently reported in *Nature*³, that inositol compounds may have actions outside cells as well as inside. Are they also involved in extracellular signalling?

Mechanisms

InsP_3 is produced in cells in response to an agonist — neurotransmitter, hormone or growth factor — binding to receptors on the cell surface (see figure), and is thought to bind to receptors on the membranes of the endoplasmic reticulum (ER) to trigger the release of stored calcium. For a compound like InsP_3 to be effective as an intracellular messenger, its concentration must be tightly regulated, so much attention is being given to the metabolic cycle that generates and removes InsP_3 (see figure). Several other higher inositol phosphates have been found in cells, including cyclic compounds and higher polyphosphates, but it is not clear that they all act as messengers. R. Irvine (Institute of Animal Physiology, Babraham) suggested that every compound detected need not have a physiological

function — some may merely be metabolic intermediates ignored by the cell. Apart from InsP_3 , InsP_4 is at present the only strong contender as a second messenger involved in calcium regulation.

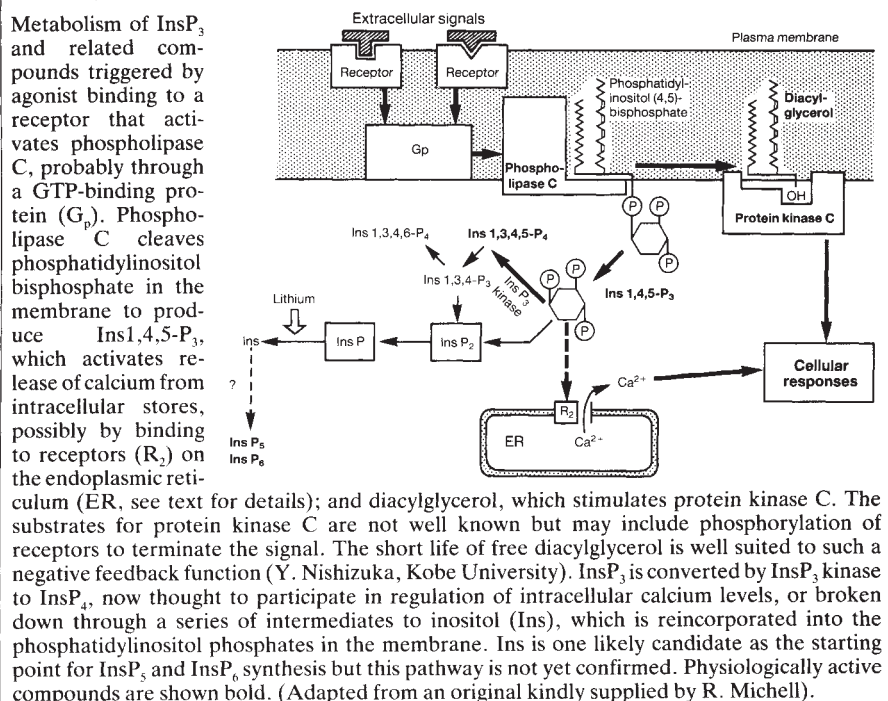
The concentration of free calcium in the cytosol can be raised either by mobilizing calcium held in intracellular pools or by promoting uptake from outside through channels in the plasma membrane. Most evidence indicates that InsP_3 is involved only in intracellular mobilization, but recently there has been an indication that, in lymphocytes, it can promote calcium uptake⁵. These experiments, however, remain inconclusive because the action of InsP_4 was not excluded (see ref. 6).

In a technically difficult experiment using patch-clamped acinar cells from

mouse lacrimal gland, Morris *et al.*² now provide evidence indicating that InsP_3 and InsP_4 act synergistically to promote calcium inflow. The rise in intracellular calcium in these cells normally stimulated by acetylcholine, which Morris *et al.* monitor by measuring a calcium-activated potassium current, can be mimicked by perfusing a cell through the patch electrode with a mixture of InsP_3 and InsP_4 . In the same cell, either InsP_3 or InsP_4 alone has no effect.

Nice though this demonstration is, problems remain and the verdict on the role of InsP_4 in calcium regulation is still open. The experiments do not exclude the possibility that the enzyme, InsP_3 kinase, that converts InsP_3 to InsP_4 (see figure) is still active in the perfused cells, which may explain why the cells tested respond to acetylcholine when only InsP_3 is present. Another explanation for the action of InsP_4 is that it could inhibit the outward calcium pump rather than stimulate calcium inflow — the net result would still appear as synergy between InsP_3 and InsP_4 to produce an increase in intracellular calcium (A. Spät, Semmelweis University, Budapest). Further doubts are raised by the discovery that some cultured cell lines, such as AR42J pancreatoma cells (J. Putney, NIH, Research Triangle) and NIH 3T3 cells (M. Wakelam, Glasgow University), seem to have no InsP_4 but are able to take up extracellular calcium. Possibly the control of calcium inflow is regulated in different ways, according to the particular needs of the cell type.

Another unsolved question is how InsP_3 , with or without InsP_4 , controls calcium mobilization. The intracellular calcium



*Inositol Lipids and Transmembrane Signalling, discussion meeting at the Royal Society, London, 2–3 December 1987. Phosphoinositides and Cell Signalling, Ciba Foundation, London, 4 December 1987. Both meetings organized by M.J. Berridge and R.H. Michell.

store is thought to be in the smooth ER and there is some evidence⁷ for InsP_3 binding sites, probably on ER membranes. But there seems to be a second pool, unaffected by InsP_3 , from which calcium is released by GTP⁸. Donald Gill (University of Maryland School of Medicine, Baltimore) now thinks that the function of this pool is to remove calcium from the cytosol. Rather than calcium being released from the second pool, he proposes that it is transferred to the InsP_3 -sensitive pool. The transfer would be promoted by GTP, which may couple the pools together, possibly under the control of InsP_4 .

It is not clear whether, in normal cells, the two pools are physically separate or simply different parts of the ER, perhaps containing different calcium-binding proteins. A new reagent, thapsigargin, may provide a probe for dissociating the release of calcium from InsP_3 production and so enable some of these ideas to be tested. Extracted from an umbelliferous plant (*Thapsia garganica*), it can discharge calcium from the agonist-sensitive pool in intact cells without stimulating inositol phosphate production (M. Hanley, Laboratory of Molecular Biology, Cambridge; O. Thastrup, University Hospital, Copenhagen).

Calcium spikes

Rapid fluctuations in intracellular calcium concentration after stimulation with an agonist have now been described in several cell types using calcium-sensitive dyes such as Fura-2 to monitor single cells. These 'calcium spikes' could reflect the coupling between the InsP_3 -sensitive and insensitive pools (M. Berridge, Cambridge University with P. Cobbold and R. Cuthbertson, Liverpool University). The period of the resulting oscillation in calcium concentration varies with cell type, from 1 second in astrocytes to 10 minutes in mouse oocytes, and spike frequency is often proportional to agonist concentration. Berridge and colleagues speculate that the spike frequency provides a digital readout of stimulus strength.

The significance of this suggestion rests in the different calcium sensitivities of various cellular functions — those with low thresholds will be recruited when the frequency is low; only at high frequencies will enough calcium be available to trigger cell division, for example. Low-frequency stimulation may produce alternating protein phosphorylation-dephosphorylation cycles, whereas at high frequencies there would be insufficient time for significant dephosphorylation. A single system could thus produce both intermittent and prolonged signals and so control different aspects of the behaviour of a cell.

What of other roles for inositol compounds, unconnected with calcium regulation? InsP_3 and probably InsP_6 seem

to be produced from inositol by a completely separate path (see figure). InsP_3 is present in considerably higher concentrations than InsP_6 but its level changes much more slowly after hormone stimulation, indicating that it is not a second messenger; both compounds could merely be involved in cellular housekeeping processes (Irvine).

But, as Hanley and colleagues reported recently³, in the mammalian central nervous system, InsP_3 and InsP_6 may be secreted from cells and have an extracellular action. Hanley *et al.* find differential regulation of the synthesis of the two compounds in different areas of the central nervous system, with InsP_3 highest in the medulla and InsP_6 in the hippocampus. Microinjections of each compound into an area of the medulla that regulates the cardiovascular system produced rapid changes in heart rate and blood pressure. Both compounds are extremely potent, 10–100 times as effective as the putative neurotransmitter glutamic acid. Injecting InsP_4 had no effect.

This is of course only very preliminary evidence that InsP_3 and InsP_6 are neuroactive. The usual battery of tests, such as demonstrations of their release and inactivation, and direct observations of an effect on membrane channels in nerve cells will be needed before they can be proclaimed new neurotransmitters or modulators. Meanwhile, they do offer potent neuropharmacological agents for investigating cardiovascular control.

In the past three years, cell-surface proteins ranging from enzymes to antigens, adhesion molecules and protozoan coat proteins have been shown to be anchored to the membrane by a phosphatidylinositol moiety linked to a diacylglycerol⁴. As previously discussed in News and Views⁹, a molecule of this type has been implicated in the long-sought second messenger for insulin. A. Saltiel (Rockefeller University, New York) now has evidence that the molecule, an inositol-phosphoglycan (IPG) with no protein attached, is released into the cytoplasm from a glycosyl-phosphoinositide precursor in the membrane. Saltiel thinks that insulin bound to its receptor stimulates an insulin-specific phospholipase C that releases the IPG into the cytosol, leaving free diacylglycerol in the membrane. His experiments indicate that the IPG could be the messenger involved in the metabolic effects of insulin stimulation but that it has no influence on initial glucose uptake. It is possible, but still untested, that at least part of the glucose transport is regulated by a protein kinase C stimulated by the diacylglycerol freed by the phospholipase action.

As the insulin-specific phospholipase C is embedded in the cell membrane, it could also cleave the phosphoinositide

anchor holding proteins on the cell surface. There are reports⁹ that heparan sulphate and perhaps alkaline phosphatase are released by this mechanism in response to insulin stimulation. Now Saltiel has preliminary evidence that lipoprotein lipase may be. Such products could also help to explain the multiple effects of insulin on cells and one wonders whether other hormones act in the same way to produce intracellular and extracellular signals simultaneously.

Intriguing possibilities

InsP_3 is now implicated in cellular processes as diverse as the light adaptation of photoreceptors (A. Fein, Marine Biological Laboratory, Woods Hole) and the contraction of smooth muscle (A. Somlyo, University of Pennsylvania, Philadelphia). Perhaps the most intriguing is the possibility that it is a mediator in pattern formation during development, as indicated by experiments described by W.B. Busa (Johns Hopkins University, Baltimore). Lithium has long been known as a potent teratogen and it has recently been shown to block the breakdown of inositol phosphate to inositol (see figure). Injecting lithium into a single cell of a 32-cell *Xenopus* embryo results in the development of tadpoles with two heads¹⁰. Busa, in collaboration with R. Gimlich (Scripps Clinic, San Diego), has now shown that this teratogenic effect can be blocked if myoinositol, a precursor for phosphoinositides, is injected at the same time as lithium. The normal tadpoles that result from this rescue treatment are a strong indication that the phosphoinositide signal cycle is involved in pattern formation, although how remains to be determined.

Finally, J. Pouyssegur (CNRS, Nice) reminded us that the inositol phosphates do not exist alone in cells but interact in complex, poorly understood ways with other intracellular signalling pathways, such as the adenylyl cyclase cascade and tyrosine kinase receptors. He demonstrated that growth factors, acting through tyrosine kinase receptors, can potentiate the production of InsP_3 but only if the phosphatidyl-inositol pathway is already active, an observation that may resolve several previously unexplained results. It looks as though research in inositol phosphate signalling is entering its high baroque era. □

1. Berridge, M.J. & Irvine, R.F. *Nature* **312**, 315–321 (1984).
2. Morris, A.P., Gallacher, D.V., Irvine, R.F. & Petersen, O.H. *Nature* **330**, 653–655 (1987).
3. Vallejo, M., Jackson, T., Lightman, S. & Hanley, M.R. *Nature* **330**, 656–658 (1987).
4. Low, M.R. & Saltiel, A. *Science* (in the press).
5. Kuno, M. & Gardner Ph. *Nature* **326**, 301–304 (1987).
6. Neher, E. *Nature* **326**, 242 (1987).
7. Worley, P.F., Baraban, J.M., Colvin, J.S. & Snyder S.H. *Nature* **325**, 159–161 (1987).
8. Gill, D., Veda T., Chueh, S.-U. & Noel, M.W. *Nature* **320**, 461–464 (1986).
9. Espinal, J. *Nature* **328**, 574–575 (1987).
10. Kao, K.R., Masui, Y. & Elinson, R.P. *Nature* **322**, 371–373 (1986).

Jennifer Altman is an Assistant Editor of Nature.

Three into two won't go

SIR—It is becoming hard to believe that after a decade of studies on gene structure and behaviour, observations of interspecific divergences and intraspecific polymorphisms are still interpreted by the constrained assumptions of the hoary old debate between selectionists and neutralists. Greenough and Harvey's recent comment¹ on Hudson, Kreitman and Aguadé's supposed discriminatory test between neutrality and selection², based on sequence variability in the *Adh* gene of *Drosophila* species, is but one of several recent discussions written in seeming ignorance of the dynamics of DNA, and which inadvertently disseminates misplaced excitement.

It is highly unlikely that there exist long tracts of sequence that genuinely fall into the category of kinetically 'complex' single-copy DNA passively picking up base substitutions and subject only to the vagaries of neutral drift or the whims of selection. DNA is an 'active' hyper-variable molecule, in a state of flux, as a consequence of several mechanisms of rearrangement that cause continual gains and losses of large and small tracts of DNA. Much of the coding and non-coding DNA is subject to one or more turnover mechanism (including gene conversion, unequal crossover, slippage and transposition). These can operate at different rates, with different biases and on different units of DNA (for some reviews see refs 3–6). In some intensely examined genes it is clear that turnover mechanisms can operate one on top of another, that they are involved with the generation and dissemination of mutations, and that they make a major contribution to observed levels of divergence and polymorphism in exons and introns (see, for example, refs 7 and 8). In all tests and discussions of molecular evolution, these internal forces of the genome have to be considered in addition to the external phenomena of drift and selection.

How does all this help us with the *Adh* data? The observations are simple to explain. Both the silent site substitutions of the exons and four kilobases of what seems to be junk DNA 5' to the gene have evolved at comparable rates in the time separating the two sibling species *D. melanogaster* and *D. sechellia* and also the two more distantly related species *D. melanogaster* and *D. pseudoobscura*³. But quite unexpectedly, given the assumptions of the neutral theory, the 5'-flanking sequences have approximately fourfold less polymorphism than the gene sequences in *D. melanogaster*. Kreitman and Aguadé¹⁰, and now Greenough and Harvey¹, are happy to conclude that selection is maintaining high levels of balanced polymorphism in the exons.

The selection hypothesis is *ad hoc* and

based on assumptions of passive DNA. Equally plausible is the *ad hoc* hypothesis that turnover mechanisms have contributed considerably to the discrepancy between interspecific divergence and intraspecific polymorphism. It is interesting that the entire 5'-flanking sequence is rich in A + T, has frequent runs of homonucleotides and many small insertions/deletions¹⁰. These features, all hallmarks of the activities of slippage, are common in flanking DNA in *Drosophila*^{10,11} and other species⁴, and generate high levels of interspecific divergence¹². Coupling this observation to the suggestion that gene conversion is operating in a domain that covers at least the 5'-flanking region, it is possible to explain relatively reduced levels of within-species polymorphism in this region.

Conversion domains can begin and end without regard for the functional requirements of the underlying sequences, and can embrace hundreds of kilobases of DNA or involve less than 10 bases^{8,13,14}. When conversion is unbiased, high levels of homogeneity ensue within a species for a given DNA region, without disturbing the rate of divergence between species. Homogeneity and conservation are not the same thing. Application of appropriate analytical methods for detecting slippage and conversion to the *Adh* gene and its flanking sequences might go some way to solving our paradox. These mechanisms could even be making a contribution to the non-random distribution of silent substitutions, clustered in the third exon⁹.

Whatever the precise causes might be for such evolutionary paradoxes, it is important to uncover the potential answers emanating from the dynamics of DNA behaviour, before squeezing the data into the restricted assumptions of classical population genetics. The same holds true for concepts of evolutionary molecular clocks¹⁵. It will be hard to escape from under the huge mathematical superstructure traditionally used to interpret the new data but a start has to be made. It is pointless for selection and neutrality to slug it out while the evidence of a third contributing force is passing them in the night. Evolution is a complex beast and we need to grasp the horns of it firmly.

GABRIEL A. DOVER

Department of Genetics,
University of Cambridge,
Cambridge CB2 3EH, UK.

1. Greenough, J. & Harvey, P.H. *Nature* **329**, 585–586 (1987).
2. Hudson, R.R., Kreitman, M.E. & Aguadé, M. *Genetics* **116**, 153–159 (1987).
3. Jeffreys, A.J., Wilson, V. & Thein, S.L. *Nature* **314**, 67–73 (1985).
4. Tautz, D., Trick, M. & Dover, G.A. *Nature* **322**, 652–656 (1986).
5. Blaisdell, B.E. *J. molec. Evol.* **21**, 278–288 (1985).
6. Levinson, G.L. & Gutman, G.A. *Molec. Biol. Evol.* **4**, 203–221 (1987).
7. Eikebush, T.H. & Burke, W.D. *J. molec. Biol.* **190**, 357–366 (1986).
8. Erhart, M.A., Piller, K. & Weaver, S. *Genetics* **115**, 511–519 (1987).
9. Schaeffer, S.W. & Aquadro, C.F. *Genetics* **117**, 61–73 (1987).
10. Kreitman, M.E. & Aguadé, M. *Genetics* **114**, 93–110 (1986).
11. Tautz, D. & Renz, M. *Nucleic Acids Res.* **12**, 4127–4138 (1984).
12. Tautz, D., Tautz, C., Webb, D.A. & Dover, G.A. *J. molec. Biol.* **195**, 525–542 (1987).
13. Shen, S.H., Slightom, J.L. & Smithies, O. *Cell* **26**, 191–203 (1981).
14. Baltimore, D. *Cell* **24**, 592–594 (1981).
15. Dover, G.A. *J. molec. Evol.* (in the press).

9. Schaeffer, S.W. & Aquadro, C.F. *Genetics* **117**, 61–73 (1987).
10. Kreitman, M.E. & Aguadé, M. *Genetics* **114**, 93–110 (1986).
11. Tautz, D. & Renz, M. *Nucleic Acids Res.* **12**, 4127–4138 (1984).
12. Tautz, D., Tautz, C., Webb, D.A. & Dover, G.A. *J. molec. Biol.* **195**, 525–542 (1987).
13. Shen, S.H., Slightom, J.L. & Smithies, O. *Cell* **26**, 191–203 (1981).
14. Baltimore, D. *Cell* **24**, 592–594 (1981).
15. Dover, G.A. *J. molec. Evol.* (in the press).

GREENOUGH AND HARVEY REPLY — Dover's suggestion that the "dynamics of DNA behaviour" provide an alternative to natural selection and neutral drift in explaining the *Adh* data of Hudson *et al.*¹ is wrong. His suggestion amounts to neither an alternative explanation nor a viable one.

Hudson *et al.* found that DNA sequence data for the *Adh* locus of *Drosophila* and its 5'-flanking region are inconsistent with the neutral model of molecular evolution: either the amount of intraspecific silent polymorphism in the gene is too high relative to polymorphism in the flanking region, or its interspecific divergence is too low¹. To account for this inconsistency, Dover suggests that slippage (a mechanism by which nucleotide sequences may gain or lose repeat units) has increased the rate of divergence in the flanking region, while conversion has maintained its low variability.

Dover's specific proposal is wrong for at least three reasons. First, divergence was determined from aligned base sequences¹, so the possibility of any effects due to slippage (which causes length mutations) was eliminated.

Second, as we reported², Hudson *et al.* can eliminate the entire class of explanations based on discrepancies in the divergence data: the data show equal rates of divergence in the silent sites of the *Adh* locus and the flanking region, which accords with the expectation that the two regions are equally unconstrained by selection (that is, they are equally free to change). Dover's assumption that the flanking region would have diverged less were it not for the intervention of "the internal forces of the genome" leads to the conclusion that the flanking region is more constrained by selection than the silent sites of the *Adh* locus. And yet, not only does the flanking region have no open reading frames, but insertion/deletion polymorphisms are tolerated in the region, and tests have failed to find any lethal-mutable loci or any large regions affecting *Adh* expression there³.

Third, again as we reported² from Hudson *et al.*¹, the problem in the *Adh* data seems to be too much variability at the *Adh* locus rather than too little in the 5'-flanking region. The proportion of nucleotide sites which are polymorphic in the 3'-flanking region is about one-third that in the *Adh* locus. Furthermore, most of the polymorphism of the *Adh* locus is

clustered in one of the three exons: the proportion of nucleotide sites that are polymorphic in this exon is three and a half times that in the other exons and between two and four times that in the introns³. Thus, Dover would need to postulate that conversion influenced both flanking regions, both introns and two exons, but not the third exon! Hudson *et al.* have a more parsimonious interpretation.

Furthermore, the theoretical foundation for some of Dover's claim is lacking. Dover offers no support at all for his claim that unbiased conversion will decrease the polymorphism at a locus but not affect its rate of divergence. Because unbiased conversion amounts to an additional mechanism of drift, one might expect that it would increase the rate of divergence between species. Until we see a formal model, we will remain unconvinced that unbiased conversion has any effect on intraspecific nucleotide polymorphism beyond that attributable to drift.

Do Dover's 'turnover mechanisms' have any potential at all to explain the *Adh* data? Slippage, unequal crossing over and transposition all generate length mutations. Thus none of them can explain the *Adh* data where aligned nucleotide sequences were used to determine divergence, and length variations were excluded from the measure of polymorphism. These mechanisms have no special significance to tests of this sort. They act simply to increase the mutation rate, and thus cannot increase divergence between species without also increasing polymorphism within species. They cannot generate discrepancies from the neutral model as long as we are careful to count the result of a single mutational event as no more than one change (this is most easily done by excluding length mutations, as Hudson *et al.* did).

Unbiased conversion is just a mechanism of drift, but biased conversion is another matter. It is natural selection at the level of the gene and, like other selective forces, can increase or decrease either polymorphism or divergence, or both. For example, the rate of divergence would be decreased if the currently fixed allele had the ability to convert other alleles. This amounts to the gene being constrained by selection: new alleles are at a disadvantage, polymorphism is reduced, and there is no departure from the neutral model. As we noted², departures can arise from the fixation of a new selectively favoured allele, whether it is by biased conversion or selection at the phenotypic level. Here polymorphism is reduced while overall divergence is unaffected.

More relevant to the excess polymorphism of the *Adh* locus is the possibility that a system of balanced, biased conversion among three or more alleles might allow an accumulation of neutral

polymorphism on each allele. This possibility is consistent with the conclusion of Hudson *et al.* that the excess variation in one exon of the *Adh* locus suggests a selectively maintained polymorphism in the area. The only viable explanation for the *Adh* data that we can construct from Dover's turnover mechanism is essentially the one proposed by Hudson *et al.*

What can we conclude about Dover's general point, that we need to consider turnover mechanisms as a force in molecular evolution which is distinct from natural selection and neutral drift? Where they are unique events, slippage, unequal crossing over and transposition are simply mechanisms of mutation. Biased conversion, or any given mutation that occurs repeatedly, is a form of natural selection at the gene level (which may or may not be opposed by selection at higher levels). Unbiased conversion is a mechanism of drift. In short, the effects of turnover mechanisms may be understood within the contexts of natural selection and neutral drift. To the extent that they are special at all, it is only because they provide a means to transfer information between members of a repeated gene family. All mutations may not be equally likely, and they may differ in their forward and backward rates, but these possibilities are already incorporated into existing population-genetics models⁴. We are reminded of Maynard Smith's concluding statement to the book edited by Dover and Flavell⁵: "people have been thinking about evolution for a long time, and . . . some of their conclusions may be worthy of attention."

There is no mysterious third force. The *Adh* data will ultimately be explained with reference to natural selection and neutral drift. We need to determine the relative roles of these two forces, and the test developed by Hudson *et al.*, which was the subject of our original report, goes a considerable way towards that goal.

JULEE A. GREENOUGH

PAUL H. HARVEY

Department of Zoology,
University of Oxford,
South Parks Road,
Oxford OX1 3PS, UK

1. Hudson, R.R., Kreitman, M.E. & Aguadé, M. *Genetics* **116**, 153-159 (1987).
2. Greenough, J.A. & Harvey, P.H. *Nature* **329**, 585-586 (1987).
3. Kreitman, M. *Oxf. Surv. evol. Biol.* **4**, 38-60 (1987).
4. Falconer, D.S. *Introduction to Quantitative Genetics* 2nd edn (Longman, London, 1981).
5. Dover, G.A. & Flavell, R.B. (eds) *Genome Evolution* (Academic, London, 1982).

Is there a role for herpesvirus in AIDS?

SIR—Latchman¹ has challenged the significance of our original observation² on herpes simplex virus (HSV)-induced human immunodeficiency virus (HIV) *trans*-activation³. He argues: first that the

observed transcriptional activation of HIV is nonspecific, as others have shown that HSV infection can activate heterologous genes; second that the activation is limited to the transfected gene and does not generally apply to endogenous genes, as a transfected, but not an endogenous, beta-globin gene can be *trans*-activated; and third that the system used may not be relevant to the regulation of expression of the proviral HIV genome, because transcription of other viral genes is significantly suppressed by HSV infection.

Although valuable, Latchman's reasoning is incomplete and we would like to add three comments. First, the *trans*-activation of HIV-LTR by HSV infection is specific in that none of the heterologous promoters were activated within the same experiment. Furthermore, in our most recent studies³, we have localized a 73-base-pair sequence from the HIV-LTR which can confer HSV responsiveness to a heterologous gene that was previously unresponsive to HSV infection. Second, the HIV genome is not an endogenous gene, therefore the differential activation of a transfected gene and endogenous genes is irrelevant for this system. There is no evidence for site-specific retroviral integration in infected cells, and thus the integration of provirus into the cellular chromatin upon either infection or transfection will be a random event. Third, Latchman states that the "studies involving cells latently infected with HIV itself will be necessary to confirm the provocative suggestion of Mosca *et al.*". We have addressed this statement directly by constructing a permanent SW480 colon carcinoma⁴ cell line containing HIV proviral DNA. This cell line has a low constitutive level of reverse-transcriptase activity which is augmented 14-fold by HSV-1, 12-fold by transfection with the combination of HSV IE175 (ICP-4) and IE110 (ICP-0) genes, and 13-fold by transfection with the HIV-1 *tat1/III* gene. These data indicate that infection with HSV is able to enhance the replication of infectious HIV.

In addition to our results, recent data of Quinn *et al.*⁵ suggest that individuals with previous herpesvirus infections are particularly susceptible to either HIV infection or disease progression. Thus, it is becoming increasingly clear that herpesviruses have a role in HIV infection.

J.D. MOSCA

D.P. BEDNARIK

N.B.K. RAJ

P.M. PITHA

The Johns Hopkins University
Oncology Center, Baltimore,
Maryland 21205, USA

1. Latchman, D.S. *Nature* **325**, 487 (1987).
2. Mosca, J.D. *et al.* *Nature* **325**, 67-70 (1987).
3. Mosca, J.D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 7408-7412 (1987).
4. Adachi, A. *et al.* *J. Virol.* **59**, 284-291 (1986).
5. Quinn, T.C. *et al.* *J. Am. med. Soc.* **257**, 2617-2621 (1987).

Molluscan extinction rates in question

SIR—The Palaeogene molluscan record of the North American Gulf Coastal Plain rivals that of the Paris Basin and provides an excellent database for studies of extinction and speciation events. Palaeocene and Eocene molluscs of this basin are documented in the catalogue of Palmer and Brann¹. Recent monographs on Early Oligocene molluscs^{2,3} provide a basis for studying faunal changes across the Eocene/Oligocene boundary. In their review, Hut *et al.* cite⁴ stepwise extinction of molluscs across the Gulf Coast Eocene/Oligocene boundary as evidence for multiple cometary impacts. But they overstate the significance of these 'stepwise' extinctions, which occur between successive formations, in view of Palaeogene extinction rates in the northern gulf as a whole.

The average molluscan extinction rate between successive formations in the northern gulf Palaeogene is 70% (on the average 50% of a formation's fauna are known only from that formation). This rate increases to 95% at group/stage boundaries. The first and third steps of the Hut *et al.* molluscan extinctions correspond respectively to the Claiborne/Jackson and Jackson/Vicksburg group boundaries. Extinction rates cited within the Jackson Group between the Moodys Branch and Yazoo Formations were 72% for gastropods and 63% for bivalves. These rates are in line with the Palaeogene interformational average.

High molluscan extinction rates occurring at group/stage boundaries in the North American Gulf Coastal Plain are associated with major marine regressions⁵ that subdivide the Palaeogene sequence. These regressions were accompanied by periods of delta progradation and increased shelf turbidity. A change in the shelf environment from a clear-water, sandy-bottom shelf to a turbid, muddy-bottom shelf adversely affected most of the molluscan species and was a probable cause of their demise.

DAVID T. DOCKERY III
Mississippi Bureau of Geology,
PO Box 5348,
Jackson, Mississippi 39216, USA

1. Palmer, K.V.W. & Brann, D.C. *Bull. Am. Paleont.* **48**, 1-1057 (1965-66).
2. Dockery, D.T. III *Mississippi Bur. Geol. Bull.* **123**, 1-261 (1982).
3. MacNeil, F.S. & Dockery, D.T. III *Mississippi Bur. Geol. Bull.* **124**, 1-415 (1984).
4. Hut, P. *et al. Nature* **329**, 118-126 (1987).
5. Dockery, D.T. III *Palaio* **1**, 582-589 (1986).

HANSEN REPLIES—Dockery is correct in that average interformational extinction rates in the Gulf Coast Palaeogene are high, but they are probably lower on average than the figures he cites. I have based my calculations on the sources cited by Dockery but I have also 'filtered' the data by including only species from the clastic

sedimentary provinces of the Gulf Coast (specimens from carbonate facies are usually poorly preserved, fossiliferous formations in the Atlantic states are patchy and represent a different biogeographic province making it difficult to correlate with the Gulf Coast), eliminating all unnamed species (such as *Nuculana* sp.), and omitting all species named from a single poorly preserved specimen or in which the sole type specimen has been lost. All these criteria reduce the number of 'short-lived' species, reduce the average extinction rate and make the species more equivalent to true biological species. These adjustments reduce the mean interformational extinction rate to around 50% (from 70% cited by Dockery). In addition, it is misleading to compare the Late Eocene extinction rates with the average rate because Palaeogene extinction rates are not uniformly high. Rather they are high during the Early Palaeocene (60-70%), possibly because of high faunal turnover in the wake of the Cretaceous-Tertiary extinctions¹, high near the Palaeocene-Eocene boundary (around 80%), low through most of the Middle Eocene (around 25%) and high approaching and during the Late Eocene (70-90%). The fact that extinctions tend to be particularly high near some group/stage/epoch boundaries is not surprising because these boundaries have been named in part precisely because of high extinctions and the resulting faunal changes.

The Late Eocene molluscan extinctions have not yet been studied in the microstratigraphic detail characteristic of the Cretaceous-Tertiary boundary, but they are 'stepwise' in that rather than a single abrupt extinction at the Eocene-Oligocene boundary (a common impression given by literature studies of this extinction²), the molluscs undergo a series of accelerated (higher than average) extinction episodes starting at the Mid-Late Eocene boundary and continuing to the Eocene-Oligocene boundary³. The immediate cause of these extinctions has been debated already³⁻⁶. Dockery's choice of local shelf turbidity as a primary control is a poor one because the molluscan extinctions tend to be selective for warm-water taxa, and not for taxa intolerant of high turbidity, and they broadly correlate with stepwise extinctions among the planktonic foraminifera from deep-sea Atlantic cores^{1,6,7}.

THOR A. HANSEN
Department of Geology,
Western Washington University,
Bellingham, Washington 98225, USA

1. Hansen, T.A. *Paleobiology* (in the press).
2. Raup, D.M. & Sepkoski, J.J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 805-805 (1984).
3. Hansen, T.A. *Palaio* **2**, 69-75 (1987).
4. Dockery, D.T. III *Palaio* **1**, 582-589 (1986).
5. Dockery, D.T. III *Palaio* (in the press).
6. Hansen, T.A. *Palaio* (in the press).
7. Hut, P. *et al. Nature* **329**, 118-126 (1987).

Cometary organics

SIR—Chyba and Sagan write¹ "Hoyle and N. Wickramasinghe compare the 3-4 μ m spectrum of comet Halley obtained at the Anglo-Australian telescope by D. Wickramasinghe and Allen with the prediction of a bacterial model. We believe that the agreement between the two provides no evidence of cometary bacteria".

When a pre-existing theory turns out to fit very well to later observations is this really "no evidence"? If so, science is nowadays being prosecuted according to principles very different from those used in the past, when such an agreement was construed as positive evidence for the theory, although whether could be considered to prove the theory was of course another question.

Chyba and Sagan ask for a more explicit statement of our "modelling procedure". Our emission curve^{2,3} was essentially calculated from the formula $A\tau(\lambda)B_{\lambda}(T)$, where A is a normalization constant adjustable to fit the quantity of material around comet Halley, $\tau(\lambda)$ is the measured emissivity of bacteria already published several times⁴, and $B_{\lambda}(T)$ the Planck function, with T given as 320 K.

Subsequent refinements to this model have taken account of a size distribution of grains and used the Mie formulae to calculate both the scattering background and the distribution of grain temperature under various conditions⁵. The correspondences shown earlier were not significantly altered by these later refinements.

Other points raised by Chyba and Sagan are based on the conception that a particular unchangeable grain model should be capable of explaining observations made at different times. Because of the sporadic activity of the comet, however, no one grain model can be expected to explain all the observations at all times. Our initial arguments applied specifically to the situation on 31 March 1986. On that occasion the observations of D. Wickramasinghe and Allen emphasized those particles that had the best ability to absorb sunlight in the visible region of the spectrum, as only particles that become heated through absorption in the main part of the solar spectrum would have been able to radiate appreciably in the 3-4 μ m region of the infrared. Such absorption would be conveniently explained quantitatively by pigments, which could vary considerably between different particles (especially between organic and inorganic particles), and which would probably change with time as the pigments were exposed progressively to solar ultraviolet light.

Chyba and Sagan say that an attempt "to explain the spectrum of comet Halley with living organisms or their products seems an extravagant departure from Occam's razor". Why? We know that some 10^{14} - 10^{15} kg of organic material is produced

every year by living organisms on the Earth. What definite observation can Chyba and Sagan cite to support their belief in the production of a specific mixture of organic materials in significant quantity from "irradiated candidate ices"?

Chyba and Sagan apply Occam's razor the wrong way round because of the strongly-held cultural prejudice that there is no life outside the Earth or, if there is, extraterrestrial life can have no connection with terrestrial life⁶. We have little objection to those who refuse to be swayed by our argument because of this cultural belief. Our objection is to its use to bolster criticisms that would otherwise be seen to have little merit.

Chyba and Sagan's most recent attempt⁷ to match the spectrum of comet Halley using an "irradiated candidate ice" along with two other arbitrarily added components reflects an even further violation of Occam's razor. Not only is their best fit considerably worse than ours, but the maximum yield of organics to be derived from their methane-clathrate model is limited to about 0.1 per cent of the water content, less than that observed in the comet by a factor of at least 100. Moreover, the requisite irradiation conditions cannot in our view have been achieved in the Solar System.

F. HOYLE

N.C. WICKRAMASINGHE

University College,
Cardiff CF1 1XL, UK

1. Chyba, C. & Sagan, C. *Nature* **329**, 208 (1987).
2. Hoyle, F. & Wickramasinghe, N.C. *Nature* **328**, 117 (1987).
3. Wickramasinghe, D.T., Hoyle, F., Wickramasinghe, N.C. & Al-Mufti, S. *Earth, Moon Planets* **36**, 295 (1986).
4. Hoyle, F., Wickramasinghe, N.C., Al-Mufti, S., Olavesen, A.H. & Wickramasinghe, D.T. *Astrophys. Space Sci.* **83**, 405 (1982).
5. Wallis, M.K., Rabizirov, R., Wickramasinghe, N.C. & Al-Mufti, S. *ESA-SP 278* (1987).
6. Hoyle, F. & Wickramasinghe, N.C. *Nature* **322**, 509 (1986).
7. Chyba, C. & Sagan, C. *Nature* **330**, 350–353 (1987).

SIR—The abundant complex organic refractory molecules revealed by the Vega/Giotto mass-spectrometer analysis of the dust of comet Halley¹ and the 3.4- μ m emission observed from space and from the ground by five groups (for example, ref. 2) are the subject of controversy. What are they and where did they come from? Have they existed in the comet since birth or were they created subsequently: Chyba and Sagan's³ challenge to the Hoyle and Wickramasinghe⁴ bacteria model of comet Halley dust is welcome because it clearly states several criteria for accepting or rejecting possible candidates for the 3.4- μ m emission. Although these criteria are useful for rejecting the bacteria model, Chyba and Sagan's alternative model has its own problems.

We agree that the major criteria for the identification of the 3.4- μ m material of comet Halley are that there must be a "plausible account of the origins of the organic material, a plausible model for the

infrared emission of this material, and a demonstration that this conjunction of material and model not only matches the 3.4- μ m spectrum". Of these three criteria, only the second is satisfied by Chyba and Sagan.

A combination of theory, observations and laboratory studies has shown that photoprocessing of interstellar dust mantles accounts for the presence of complex organic material as a major component of the dust^{5,6} and of comets considered as aggregated interstellar dust⁷. Our studies of the refractory materials produced under laboratory conditions simulating the interstellar processing give a reasonable fit to the 3.4- μ m interstellar dust absorption to the galactic centre and some provide a better fit to the comet Halley emission feature than does the material referred to by Chyba and Sagan³. No published laboratory residue spectrum provides a perfect fit to the Halley dust emission, although that of Chyba and Sagan is closer than that of Hoyle and Wickramasinghe.

This proves only that the 3.4- μ m emission comes from some mixture of organic molecules containing CH₂ and CH₃ groups. So we agree that the mere existence of a 3.4- μ m emission does not justify the bacteria model. But we totally disagree with the alternative suggestion that the "radiation processing by comet Halley" and the resulting "residue of irradiated lowoccupancy methane ice clathrate" can provide the abundant organic refractory components seen in the comet dust. First, from what we now know of comet Halley there is little to justify the assumption that a methane ice clathrate has any relevance to the comet nucleus. Second, the radiation processing of comet Halley during its 4.6×10^9 years of exposure to cosmic rays (no other processing is relevant once the comet formed) can have only produced significant effects to the interstellar dust (already photoprocessed) to a depth of about five metres or less⁸. Thus the modified outer layers have long since been shed by comet Halley since its first recorded apparition some 2,000 years ago. On the other hand the organic refractory interstellar dust is still evident because it was homogeneously incorporated throughout the nucleus.

With regard to the physical requirements for the emission we agree with Chyba and Sagan that the choice of a single temperature to characterize both the continuum and emission feature is incorrect and that the 3.4- μ m emission must be due to much hotter particles⁹ ($T \geq 480$ K at 1 AU) than those at a little above black-body temperatures ($T \approx 320$ K at 1 AU) which produce the continuum¹⁰. We also agree that particles with a radius of less than about 1.0 nm will be much hotter than large ones (or black bodies)¹¹ and that there must be

organic particles in this size range. But Chyba and Sagan nowhere state why their proposed formation of organics in comets should provide a source of such small particles nor why, by apparent coincidence, the 10- μ m emission should also require the presence¹² of such small particles. It behoves an acceptable model to account for the simultaneous presence of hot silicate and organic particles. This simultaneous presence is automatically predicted by the grains of silicate core-organic refractory mantle interstellar dust that constitute the comet dust. The individual grains are predictably less than about 1.0- μ m in radius and fluffy aggregates¹³ of such particles up to a certain size, limited by an optical depth less than unity, will be as hot as their individual components. Larger aggregates will approach black-body temperatures.

Finally, we find it curious that although Chyba and Sagan "strongly suggest the presence of organic-coated grains in the Halley coma", they do not note that this is typical of interstellar dust with silicate cores and organic refractory mantles.

Some time ago one of us showed that the predictions for many of the observed properties of comet Halley follow inevitably from the interstellar dust model¹³. It now appears that still another observation — the simultaneous 3.4- μ m and 10- μ m emission by superheated particles — is predicted by the presence of small dust grains as individuals and in fluffy clumps. A possible additional source of 10- μ m emission, not directly associated with 3.4- μ m emission, could be the very small bare interstellar silicate particles which were trapped in the original outer icy grain mantles during comet aggregation and released on evaporation along with H₂O and other volatile constituents. These particles are too small to contribute to the scattered light continuum and have been referred to as a "fine mist"¹⁴.

Chyba and Sagan have recently published an extended account of their case¹⁵.

J. MAYO GREENBERG

NANSHENG ZHAO

Astrophysics Laboratory,
Leiden University,
2300 RA Leiden, The Netherlands

1. Kissal, J. *et al.* *Nature* **321** 280, 336–337 (1986).
2. Baas, F., Geballe, T.R. & Walther, D.M. *Astrophys. J.* **311**, L97–101 (1986).
3. Chyba, C. & Sagan, C. *Nature* **329**, 208 (1987).
4. Hoyle, F. & Wickramasinghe, N.C. *Nature* **328**, 117 (1987).
5. Schutte, W. & Greenberg, J.M. in *Light on Dark Matter* (ed. Israel, F.F.) 229–232 (Reidel, Dordrecht, 1986).
6. Greenberg, J.M. *Scient. Am.* **250**, 124–135 (1984).
7. Greenberg, J.M. in *Comets* (ed. Wilkening, L.L.) 131–163 (Univ. Arizona Press, 1982).
8. Greenberg, J.M. & Grim, R. in *Proc. ESA SP-250*, 255–263 (1986).
9. Greenberg, J.M. & Zhao, Nansheng in *Fluffy Particles* Leiden Workshop 1987 (in the press).
10. Sagdeev, R.Z. *et al.* *Nature* **321**, 259–262 (1986).
11. Hanner, M.S. in *Cometary Exploration II* (ed. Gombosi, T.I.) 1–22 (Hung. Acad. Sci., 1982).
12. Bregman, J.D. *et al.* *Astr. Astrophys.* (in the press).
13. Greenberg, J.M. *Nature* **321**, 385 (1986).
14. Greenberg, J.M. *Adv. Space Res.* **4** (9), 211–212 (1984).
15. Chyba, C. & Sagan, C. *Nature* **330**, 350–353 (1987).

Propheteering business

Martin Gardner

Bare-Faced Messiah: The True Story of L. Ron Hubbard. By Russell Miller. *Michael Joseph, London: 1987. Pp. 390. £12.95. To be published in the United States by Henry Holt.*

L. Ron Hubbard: Messiah or Madman? By Bent Corydon and L. Ron Hubbard, Jr. *Lyle Stuart, 120 Enterprise Avenue, Secaucus, New Jersey 07094: 1987. Pp. 402. \$20.*

FOR 35 years I believed that L. Ron Hubbard, founder of Scientology, was no more than a writer of mediocre fiction who, lusting for power and money, became one of the world's most successful mountebanks. Russell Miller's admirable, meticulously documented biography has persuaded me otherwise. Hubbard was a deeply disturbed man — a pathological liar who steadily deteriorated from a charming rogue into a paranoid ego-maniac "unable to distinguish", as Miller puts it, "between fact and his own fantastic fiction".

Almost everything Ron ever said about himself was false. He was never a swashbuckling explorer or distinguished naval officer. Although he claimed to be a physicist, his knowledge of science was negligible. His father, a lieutenant-commander in the US Navy, had hoped his son would pursue a similar career, but near-sightedness kept Ron out of Annapolis. His only education was in the engineering school of George Washington University where he dropped out after two years of dismal grades.

During the ten years preceding the Second World War, Ron became one of the nation's most industrious contributors to western, mystery and adventure pulp fiction. His four years in the wartime Navy are summed up in a fitness report saying he "lacked essential qualities of leadership . . . not considered qualified for command or promotion". The closest he came to combat was while in command of a submarine chaser. On its shakedown cruise Hubbard mistook a magnetic deposit for submarines, and his battle against the non-existent enemy cost him his command. Assigned to a ship on its way to a war zone, he at once applied for and obtained transfer to a school at Princeton. "Far from being a hero", Miller concludes, Hubbard was a "malingering coward who had done his best to avoid seeing action".

Ron's record as husband, father and bigamist was even more deplorable. He deserted his first wife, Polly Grubb, and

their two children to marry Sara Northrop. They met in the 'temple' of Jack Parsons, an addlepated Californian chemical-propulsion expert who secretly practised black magic. A firm believer in witchcraft, Parsons had become a devoted disciple of England's 'Beast 666', the notorious satanist Aleister Crowley. Hubbard,

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Sensitive salad — Hubbard goes to work on a tomato.

Parsons informed his 'Blessed Father', was a kindred spirit eager to assist in blasphemous rituals. Ron left the temple with Parsons's young mistress Sara. A few years later, Parsons blew himself to Hades by accidentally dropping a phial of nitroglycerin. His mother, hearing the news, killed herself with sleeping pills.

Hubbard's bigamous marriage to Sara, which produced a daughter, lasted five years. Polly of course divorced him. In Sara's later divorce action he was said to be "hopelessly insane". In 1952, aged 41, Hubbard married Mary Sue Whipp, 19, by whom he had four more offspring. When their son Quentin committed suicide, Hubbard's only reaction was one of fury.

It was while Ron was hacking out science fiction that he conceived of

dianetics. In this hilarious parody of psychoanalysis, ills are said to spring from 'engrams' recorded on an embryo's brain by what it overhears even before it develops ears. After engrams have been erased by 'auditing', one becomes a 'clear', with perfect memory and robust health. The new science was released to the world in a rousing article by Hubbard in *Astounding Science Fiction*, quickly followed by his book *Dianetics*. The ludicrous therapy exploded into a national craze.

Hubbard discovered that a crude lie detector he called an E-meter was a valuable auditing tool. With its aid, 'pre-clears' were soon recalling not only birth traumas but previous lives. Ron saw at once that by combining dianetics with reincarnation he could fabricate an exotic 'religion' capable of raking in millions of tax-free dollars. The Church of Scientology began with recruitment of naive youngsters who really believed Ron had found a 'bridge' to transcendent realities that would transform the world. The bridge grew more baroque as dozens of new books by Ron improved the 'tech' (auditing technology) and elaborated the mythology.

Hubbard began to annoy the FBI with wild reports of Communist persecution. The Bureau considered him psychotic. On the estate of the Maharajah of Jaipur in Sussex, which Hubbard had purchased, he conducted plant research to "reform the world's food supply". E-meters convinced him that tomatoes scream when sliced. Mary Sue learned she had once been D. H. Lawrence. Ron revealed he had written *The Prince* but "that son-of-a-bitch Machiavelli stole it from me". During an incarnation on another planet, Hubbard said, he had managed a factory that made steel humanoids.

Harassed by real and fancied enemies, Hubbard sought escape by buying three ships. From his flagship *Apollo*, resplendent in officer's uniform, he became the naval commander his father had wanted him to be. Pubescent little girls in mini-skirts and high-heeled boots transmitted the 'Commodore's' orders, lit his cigarettes, washed his hair, dressed and undressed him. Hubbard's temper tantrums, abusive language and dictatorial behaviour grew more perverse. A young woman aboard shot herself to death.

For years Hubbard's Sea Org, as he called his fleet, wandered around the eastern Atlantic, its Commodore convinced

that Nazis and Reds were chasing him. Crude slapstick efforts were made to take over Rhodesia and Morocco. A proof that Hubbard — now fat, flabby-faced and impotent, with hair to his shoulders and rotting teeth — had come to believe his mythology is that his crew wasted months searching for treasures he recalled having buried in earlier incarnations.

Operation Snow White was Hubbard's secret plan to infiltrate federal offices and steal material related to his church. The plan was carried out by loyal scientologists firmly convinced they were obeying higher laws. In 1977 FBI agents broke into two of Hubbard's headquarters and carted off 48,149 documents. Mary Sue and eight others were convicted, fined and incarcerated.

Ron escaped punishment because no one could find him. For a while he hid out in a Nevada desert where, with no previous experience, he struggled to make fantasy films. The Church was ordered to allocate "unlimited funds" for a campaign to get their leader a Nobel prize. Mary Sue was released after a year in prison, a pathetic woman who to this day, perhaps out of fear, will not talk about the man who abandoned her.

Hubbard seems to have spent most of his hidden years churning out more science fiction. St Martin's published his 800-page *Battlefield Earth*. Ten even more worthless novels have since been issued by the Church and lavishly advertised. After promoting himself to Admiral, his red hair now white, Hubbard apparently died in Creston, California, on January 24, 1986. He had been living in a motor home parked on a ranch he had recently bought.

This is only the barest sketch of the sordid details in Miller's incredible account. An even more savaging life of Hubbard, by his oldest son in collaboration with another scientologist who "blew the Org", has just been published in the United States. It is carelessly written, poorly organized and documented, with neither bibliography nor index. It does, however, have a useful glossary of Scientology's major neologisms, and a chilling chapter on the cult's heartless efforts to destroy Paulette Cooper for writing *The Scandal of Scientology*.

I finished the two books with double amazement. How could a man this crazy have lived to 74 without being committed? How could a science-fiction cult, with such preposterous doctrines and evil morals, continue to flourish? Idiotic religions, I suppose, like old soldiers never die, and centuries can pass before they finally fade. □

Martin Gardner, a science writer, is author of Science: Good, Bad, and Bogus, The Whys of a Philosophical Scrivener and many other books on science, pseudoscience and mathematics.

Making the most of models

John R. Krebs

The Latest on the Best: Essays on Evolution and Optimality. Edited by John Dupré. MIT Press: 1987. Pp. 359. \$27.50, £24.75.

THERE is a remarkable and persistent division, which sometimes appears emotional rather than rational in origin, between those for whom adaptation arising through Darwinian natural selection is the blindingly simple and utterly compelling explanation of much of the diversity of life, and those for whom the concept of adaptation is tautological, teleological, untestable or demonstrably false. The division is not just between professional biologists, but occurs regularly in the serious media, amongst philosophers and in the writings of certain cosmologists, physicists and mathematicians who dabble in evolutionary biology.

One facet of this general disagreement about adaptation is the continuing debate about the role of optimality modelling in biology. Dupré's edited collection of essays brings together different views on the subject. The overall message from the book is encouraging in the sense that those who have in the past been strident critics of optimality, such as the population geneticist Richard Lewontin and the philosopher Philip Kitcher, present a gentler attack and even slightly encouraging prognosis; while users of optimality models, such as anthropologist Eric Alden Smith, ecologist John M. Emlen and psychologist John Staddon, are scrupulous in emphasizing the difficulties inherent in their work.

The book does not, of course, come up with simple solutions, but I found that nearly every chapter made useful points even if they were, as in the case of John Maynard Smith's customarily lucid essay, mainly directed at clearing up the mess left by another contributor. Rather than try to summarize the contents of particular chapters, I will try to sketch some of the general arguments that run through the book.

First, why do biologists use optimality models? The premise upon which such models are used to study adaptation is roughly this: natural selection, because it is a competitive and cumulative process, gradually tends, through many generations, to produce better and better solutions to the problems posed by an organism's environment (surviving in an arid desert, avoiding predators and so on). The result is that organisms have the appearance of being exquisitely well 'designed' to fit their environment. The

existence of design was so compelling to William Paley in 1828 that for him it was a powerful argument for a purposeful Creator, but post-Darwinian biologists agree that apparent design in nature arises through the process of natural selection with no foresight (Richard Dawkins's Blind Watchmaker). When evolutionary biologists talk about design or function, they are simply using shorthand for products of natural selection.

Optimality models are used as a tool for studying design; they are a way of asking what kind of properties organisms or bits of organisms should have if they have been selected for in a particular way. The models provide a benchmark against which observed features (structures, behaviours, physiological properties and so on) can be compared. The point of making the comparison with specific models is not necessarily that one expects to find features that conform precisely to hypothetical optima, but that the models might help to evaluate whether or not the features in question are adaptations, and if they are adaptations, what they are for.

So much for the pros, what about the arguments against? These are clearly summarized in Lewontin's chapter. First, there are plenty of reasons, including historical accident, and genetic and developmental constraints, why natural selection should not produce optimal design, or anything like it. Secondly, in building an optimization model of design, one has two kinds of parameter: the *currency* (or criterion scale as Lewontin calls it) by which performance is judged and the *constraints* on performance. Without any independent way of checking on either of these the way is open for all kinds of adjustment to get a particular model to fit. If, on the other hand, all the constraints and currencies were known before building the model, the evolutionary process would be fully understood and hence the point of optimization modelling lost. Thirdly, most optimization models involve treating design of one part of an organism at a time (for example its mating behaviour, the structure of its bones or its feeding strategy); but because all adaptations are inter-related (the way an animal feeds may depend on its mating habits and vice versa), atomization of that sort is logically flawed and one can do no more than say that organisms are an adapted whole.

These last two points are telling and seem to be accepted by those using optimality models. The difficulty of too many unknowns is a genuine one, though advocates of optimization would argue that a combination of guesswork and knowledge of the biology of the system under study allows the modeller to feel around in the landscape of constraints and currencies and make some forward progress. The question of whether or not adaptation can be atomized seems to be an empirical one

that can only be resolved by carrying through optimization studies.

The first of Lewontin's criticisms is more problematical: for those who spend their lives as biologists because they are filled with awe by the beauty of adaptation it is a tough message to be told that such things are almost impossible because of constraints of history, genetics and development. It is tempting to conclude that the argument that design is hopelessly constrained must be wrong and the facts right. However, a counter-argument to this is that the enthusiastic observer's attention is drawn to the few cases of good design and is blind to the many cases of lack of adaptation. This brings us straight back to optimality modelling as a way of finding out whether or not a particular trait is a design feature.

Given that adaptation is worth trying to understand — after all, without the concept of function, biology would be just a mass of descriptive detail (imagine a list of the components of the mammalian immune system with no attempt to say what the system does) — optimality models have a role to play as research tools. The book does not come to a consensus about this role. At the very least optimality models have forced biologists to think about the epistemological status of the concept of design; at the most, they may provide us with a coherent general framework for the study of adaptation. □

John R. Krebs is a Lecturer in Zoology at the Edward Grey Institute of Field Ornithology, University of Oxford, Oxford OX1 3PS, UK.

Rubber economy

AFTER 74 years of publication, the *CRC Handbook of Chemistry and Physics* (the 'Rubber' book) appears in every technical library worth its salt, and is an essential facility in chemistry and physics laboratories. Now the first softcover edition of this 'bible' has been published*: though with 1,800 pages it is slightly slimmed down from the hardback classic, at \$29.95 it probably gives more scientific information per unit price than any other book in print, and is aimed at a mass student market. Although the mathematics section has been pruned to a few pages, those on elements and inorganic compounds and organic chemistry retain comprehensive coverage along with sections on general chemical and physical constants. The philosophy has been to present core data that will require revision every three or four years.

The *Handbook* remains the best source of elementary data (melting point, boiling point etc.) on chemical compounds but it

* *CRC Handbook of Chemistry and Physics*, Student Edition, published by CRC Press price \$29.95. Available in Europe, from 5 April, from Wolfe Publishing, London, price £19.95.

Indian wisdom

Robert Temple

History of Science and Technology in Ancient India: The Beginnings. By Debiprasad Chattopadhyaya with appendices by various colleagues. *Firma KLM Private Ltd, 257-B B.B. Ganguly Street, Calcutta 700-012, India: 1986. Pp. 556. Rs 325.*

THE history of Indian science has not been well documented and this fascinating book does much to remedy the situation. It deals largely with what happened before 500 BC, and its contents will be highly unexpected to most readers. The book is primarily written for an Indian audience and the author bravely attacks "Hindu revivalism" and the obscurantist doctrines of the religious fanatics of India, pointing out that even many modern Indian scientists are tainted by it. This has resulted in a widespread unwillingness among devout Hindus to acknowledge indebtedness to the Indus Valley civilization, which preceded the Aryan invasion of India and which collapsed about 1500 BC. Chattopadhyaya takes this bull firmly by the horns, demonstrating conclusively that the mathematics and geometry preserved in the Hindu *Sulvasutras* can only be Harappan (Indus Valley civilization) in origin.

The book will be tough going for those with no background in Indian studies, as the references to the Vedas and countless other matters taken for granted by an

has been slow to incorporate the new spectroscopic fingerprints of chemistry — indeed CRC publish another specialist handbook in these areas. Remarkably, the softcover edition omits such limited IR and NMR information that appears in the main handbook. Ultraviolet data are sparse and Mössbauer parameters do not appear, though X-ray and atomic line spectra of elements retain wide coverage. The recently revised *Tables of Physical and Chemical Constants* (the old Kaye and Laby) published by Longman is more informative in these areas and more imaginative in presentation of data, though of course much less comprehensive on thermodynamic matters. Surprisingly there are two separate and not totally consistent tables of elemental binding energies without cross reference, while chemical shifts in binding energies get no mention. Group theory tables are another notable omission.

For all its great strengths, the full edition of the *CRC Handbook* is a curiously unbalanced data selection for the overall reference needs of contemporary chemistry and physics students. For this reason the new edition fulfils a more limited role than its title might suggest. J.A.D. Matthew

Indian readership will be obscure or even incomprehensible. But this volume is nevertheless a masterly contribution to the history of science.

It is, for example, astonishing what the author reveals of the complexity of the brick technology of the Indus Valley civilization. One of the mathematical preoccupations of that civilization was to make widely varying constructions of bricks — of great size — having identical areas and volumes. The Indus people were thus concerned with attempting to square the circle, knew the Pythagorean theorem (as the Babylonians also did), had an accurate value of the square root of 2 to five decimal places, and indeed worked with the decimal system. They were fond of problems such as "construct a square whose area is three times the area of a given square" or "transform a square into a rectangle of the same area" or "construct a triangle whose area is equal to that of a given square". This information was preserved by the later Hindus for superstitious reasons connected with the awe in which they held baked bricks (which they were incapable of making) and to which they attributed a mystical or divine power.

Indus Valley astronomy is less fully treated because less evidence survives, but it was fairly advanced, and ingenious analysis of one datum has yielded a precise date of 2357 BC for an observation. Failure of scholars to decipher the Indus script makes various speculations on Indus astronomy controversial, though many surviving short inscriptions seem to refer to stars (represented by a fish sign).

For 1,000 years, between 1500 and 500 BC, India had no form of writing. Most science disappeared. But it was during this period that the early Hindus developed linguistic science to its highest pitch in world history, a story also told in the book. The motivation was the need to preserve immensely long texts orally without so much as a syllable out of place. Stemming from this extraordinary feat of linguistics came the foundation of phonetics, etymology, semantics, and the most complicated grammar in the world, that of classical Sanskrit. Also discussed at length is the ancient Hindu concept of *rita*, which seems to be the world's earliest formulation of the principle of natural law, and which in the Vedic period in India transcended all wishes of the gods.

A sequel volume by another author dealing with the history of Indian science of subsequent epochs will appear in due course. We can hope it will match Chattopadhyaya's book, which will be a fundamental reference work for decades to come. □

Robert Temple, Hill House, Sutton Mallet, Bridgwater, Somerset TA7 9AW, UK, is a historian and science writer with a particular interest in the early science of China and India.

Willis redeivus

Vernon Heywood

The Plant-Book: A Portable Dictionary of the Higher Plants. By D.J. Mabberley. Cambridge University Press: 1987. Pp. 706. Flexi-cover £20, \$34.50.

J.C. WILLIS is one of the select company of botanists, like Hutchinson, Engler and Strasburger, who have become eponymous with a widely used publication. In the case of Willis it is his *Dictionary of the Flowering Plants and Ferns*, first published in two volumes under another title, over 80 years ago. Although Willis's greatest achievement was to come later during his 15 years as Director of the Royal Botanic Garden, Peradeniya, Sri Lanka, where his advancement of the Garden's scientific programme led to the establishment of the country's Department of Agriculture, it is for his *Dictionary* that he is best known.

The *Dictionary*, at least in its later editions, was a near-encyclopaedic vademecum of all the families and the important genera of flowering plants and ferns that the student was likely to come across, together with explanations of botanical terminology and a summary of information about their scientific and economic value. As such it was one of the most useful botanical compendia ever published and was used by generations of students and professionals — "a familiar and cherished tool" as Sir George Taylor was later to describe it. The sixth and last edition was produced in 1931, and was reprinted five times. But in 1951, in view of the vast increase in the amount of information about plant systematics that had accumulated, it was decided to invite Mr H.K. Airy Shaw of Kew to produce a fresh edition which appeared in 1966.

The Airy Shaw revision differed from previous editions in that it confined the entries to strictly taxonomic matters and most of the information on economic uses, terminology, and common and vernacular names was omitted. Subsequently some of this information was partially updated and brought together in a companion volume *A Dictionary of Useful and Everyday Plants* (1974); that book was compiled by F.N. Howes and was seen through the press after his death by David Mabberley, the author of *The Plant-Book*.

So far, the history of this fascinating work. But how does the new book differ from the Airy Shaw edition of Willis's *Dictionary*? And will we need to use both (or all three)? I am happy to say that *The Plant-Book* is quite distinctive and will prove to be as indispensable a tool as its predecessors. It nicely complements the *Dictionary*, and although there is some considerable overlap Mabberley's work is usually to be preferred because it is more

up-to-date and accurate.

The Plant-Book attempts to cover all currently accepted generic and family names and the commonly used English names of flowering plants and ferns. Accordingly most of the generic synonyms given in Airy Shaw/Willis are omitted. On the other hand, the generic entries give the family attribution, numbers of included species, distribution and, where appropriate, information on botanical, medicinal, agricultural, horticultural or other economic uses. An abbreviated reference is given to recent generic monographs. Useful comments are included on family relationships, especially where there is dissent from the viewpoint expressed in Cronquist's *An Integrated System of Classification of Flowering Plants*, the arrangement of which is otherwise followed by Mabberley for the angiosperm families. All this is contained within a volume of 700 pages, which also includes an acknowledgement of sources (some 150 books and about 500 periodicals).

The result, as far as I can judge on the basis of random sampling and checking, is remarkably good. There is a fair sprinkling of errors or misinterpretations, but this is inevitable and on the whole the book inspires the confidence which is most important in work of this kind.

The book is well printed and bound in a plastic case. It is indeed portable as the subtitle indicates and I predict that it will be commonly carried by students on their visits around botanic gardens as was Willis in my day. But more use of it will be made as a constant reference work by botanists, gardeners, agriculturists and others. Willis finished the preparation of his original book on the voyage out to Ceylon (Sri Lanka) — the island of Serendipity of the ancient mariners. What a marvellous serendipity this book is! □

Vernon Heywood, a Professor in the Department of Botany, University of Reading, Whiteknights, Reading RG6 2AS, UK, is currently on secondment to The International Union for Conservation of Nature and Natural Resources.

Radical thinking

John F. Ward

The Chemical Basis of Radiation Biology. By C. von Sonntag. Taylor & Francis: 1987. Pp.515. £49, \$98.

RESEARCH into the biological effects of ionizing radiation has commonly been concerned with the physics of energy deposition rather than the chemical processes involved. These processes involve the reactions of the free radicals, produced as a result of ionizations, with biomolecules. Recently there appears to be a greater acceptance of the contributions that chemistry can make — for instance, definition of the types of molecular damage caused within a cell by radiation (that is, the products of radical damage), and elucidation of the mechanisms of molecules known to modulate the action of radiation on cellular systems (by interference with radical processes).

The publication of this book is therefore timely. It is the first comprehensive account of the radiation chemistry of biological molecules. Beginning with a brief description of the techniques of radiation chemistry, it leads into a survey of the reactions of the free radicals which are produced by water radiolysis. The heart of the book, and its greatest value, lies in the detailed descriptions of the radiation chemistry of the biomolecules: nucleic acids, sulphur compounds, carbohydrates and proteins. Von Sonntag presents a detailed survey of these topics, drawing his material from an impressively broad literature base. His discussions are generally well argued and balanced, and,

importantly, he points out where there are discrepancies among the findings of different groups. Where he considers that presentation of the full radiation chemical background to any topic is not within the purview of his book, he provides adequate references for further reading.

Some may quibble with the author's choice of emphasis or selection of references in particular sections, and with his language usage in some places (he mentions his apprehensions in these regards in the preface). In my view, however, the presentation and the language give the text von Sonntag's personal imprint and do not detract from its value. Moreover, the book has the usual advantages of a single-author work in its continuity and consistency from chapter to chapter.

In general, the contents appear to be error-free, despite the apparent mistake on the front cover: there, the upper structure is probably meant to be the glutathionyl radical, but it lacks two methylene groups (surely licence taken by the cover designer in attempting to attain balance).

Radiation chemists, particularly those wishing to move into radiation biology, will value this book for its comprehensive survey of the field and will use it as a reference source. Radiation biologists are more likely to use it as a text, though it will present a challenge to those without a good background in chemistry. In addition, the book should be of great use to those interested in oxygen radical damage who may be unaware of the wealth of information on the reactions of these radicals in the radiation literature. □

John F. Ward is Chief of the Division of Radiation Biology, Department of Radiology, University of California at San Diego, La Jolla, California 92093, USA.

Wrong science and right science

The attack by T. Theoharis and M. Psimopoulos on philosophers of science continues to provoke correspondence. This week's Commentary pages are turned over to the matter.*

SIR—In their article, Theoharis and Psimopoulos claim that recent philosophers such as Popper, Lakatos and Kuhn (1) have betrayed the old ideals of objectivity and truth and (2) have thus played important roles in putting British science in its endangered plight.

They present not a shred of evidence for the second claim. But in the unlikely event that those who influence the purse strings have been mugging up on their recent philosophy of science, perhaps some of them have noticed passages like the following: "Science, as such, has no social responsibility. . . . [I]t is society that has a responsibility — that of maintaining the apolitical, detached scientific tradition and allowing science to search for truth in a way determined purely by its inner life."

This is from Imre Lakatos; and as for Popper, surely very few can join Theoharis and Psimopoulos in having read him without being impressed by his respect for science, that "magnificent adventure of the human spirit" and its "miraculous

success". Both Popper and Lakatos claimed to defeat the sceptic by accepting what is cogent in his argument and showing that objective (but no longer utopian) standards of scientific progress can still be developed. If their claims ultimately fail, it is for rather complicated and arguable reasons that are hardly likely to be the talk of Whitehall.

Of course Theoharis and Psimopoulos are right to point out that British science needs to fight back if its alarming predicament is not to worsen still further. But matters are not likely to be improved by attacking people who in all probability have no influence and who are anyway on their own side; nor by basing that attack on unsupported claims, caricatures of their supposed "opponents" and elementary misunderstandings.

JOHN WORRALL

*Department of Philosophy, Logic
and Scientific Method,
London School of Economics,
London WC2A 2AE, UK*

SIR—Theoharis and Psimopoulos suggest a simple, albeit naive, solution for the deep and widespread malaise of British science — the putting forth of adequate definitions of such fundamental concepts as objectivity, truth, rationality and the scientific method. But they fail to mention that this is exactly what Karl Popper, Imre Lakatos, Thomas Kuhn and Paul Feyerabend attempt. They also ignore the fact that some of the greatest minds of our century, who certainly do not qualify as "betrayers of the truth", have expressed similar opinions.

Thus Werner Heisenberg said: "The hope that the new experiments will lead us back to objective events in time and space, is about as well founded as the hope of discovering the end of the world in the unexplored regions of the Antarctic"¹. According to Max Planck, "it is impossible to separate the law that we are seeking to discover from the methods that are being used to bring about the discovery"². And Albert Einstein wrote, "The fact that in science we have to be content with an incomplete picture of the physical universe is not due to the nature of the universe itself but rather to us"³.

The authors also list three actual

dangers not only to science but to society in general. I suggest that science presents another, and in my opinion a more serious danger to society, namely elimination of God in the name of the objective truth. For example, in their book *The Third Millennium*, Stableford and Langford foresee no place for religion or God in their brave new world. Many scientists would argue that their activities have nothing to do with those issues, but for some unexplained reason a majority of them are stern atheists. At the same time, lay nonbelievers point to the achievements of science and technology as a proof for the nonexistence of God. They are convinced that scientists must know something that laymen do not; otherwise, they argue, the proportion of nonbelievers to believers in these two populations should be the same, which is not the case. Although the danger of living in a world without God is obvious to the believers, it may not be so apparent to nonbelievers.

BOGUSLAW LIPINSKI

*H.S. Research Laboratory,
1101 Beacon Street,
Boston, Massachusetts 02146, USA*

1. Heisenberg, W. *Philosophical Problems of Quantum Physics* (Ox Bow Press, Woodbridge, 1979).
2. Planck, M. *Where Science is Going?* (Ox Bow Press, Woodbridge, 1981).
3. Einstein, A. *Ideas and Opinions* (Crown, New York, 1954).
4. Stableford, B. & Langford, D. *The Third Millennium. A History of the World: AD 2000–3000* (Knopf, New York, 1985).

SIR—Theoharis and Psimopoulos comment on the ignorance of scientists about their vocation and the effect that has had, of late, on funding for research.

What I do not understand is why the two small volumes written by John Ziman have had so little influence on those worried about this problem. First in *Public Knowledge*, then in *Reliable Knowledge* (Cambridge University Press, 1968 and 1978), Ziman sets out in clear language just what distinguishes science from other scholarly pursuits and from the professions. Other scholarly disciplines do not accumulate theory and build upon it in the same way as science. On the other hand, the professions which apply scientific knowledge, such as medicine and engineering, focus on solving individual practical problems (however complex some of them may be); the art of problem-solving peculiar to each of these is an essential part of the training these professionals receive.

Ziman says that what science provides, at bottom, is best described by the title of his later book: reliable knowledge. The value of reliable information to anyone who makes decisions can obviously be very great, as is shown by the recent scandal in the United States over 'insider trading' among bankers.

Clearly, the scientific research supported by societies through their governments should be selected partly on the basis of what is likely to be important for decisions in the future. Just as clearly, that should be broadly based, both because we cannot predict all the knowledge that will follow from a particular line of investigation and because we cannot predict what kinds of decisions we will have to make some years hence. In large measure we have to argue that research of a general nature supported on general principles has proved valuable in the present, and that the past is a better predictor of the future than any other we have. Yet it is hard to argue that in a society with declining income — such as Britain — this generalized investment in the future should receive an increasing fraction of that income. The need is to tackle the difficult task of assuring that basic research does not get cut more than its share, and that the declining remainder is spent as wisely as possible.

JAMES D. WILSON

*Monsanto Chemical Company,
800 N. Lindbergh Boulevard,
St Louis, Missouri 63167, USA*

*"Where science has gone wrong" *Nature* 329, 595–598 (1987). Previous correspondence appeared in *Nature* 330, 308 and 689–690 (1987).

SIR—The diagnosis by Theoharis and Psimopoulos of the crisis in science is persuasive. It is less easy to see how the erroneous and harmful ideas which, they say, sabotage the scientific method are to be replaced by adequate definitions of objectivity, truth and rationality. There are three immediate difficulties. The first is that of rebutting the charge of vested interest; if 'objective' arguments as to the value of truth actually spring from an urge to preserve jobs, then truth itself is compromised. The second difficulty is posed by Gödel's theorem; it is vain to hope that science will ever provide 'the truth, the whole truth, and nothing but the truth', for every axiomatic system must contain propositions that are undecidable¹. The third difficulty is the problem of self-reference; how may it be possible to provide a scientific justification for doing science?

For the resolution of these difficulties we must look to areas of scholarly endeavour which lie outside science. Theoharis and Psimopoulos have already noted that a meaningful definition of truth needs observations that are context-transcendent. The problems of objectivity, completeness, and self-reference can only be resolved by invoking a value-system that is context- (that is, science-) transcendent. Although this takes us out of science, it takes us into the company of those who are skilled in manipulating and exploring theories of the transcendent — that is, theologians. Theoharis and Psimopoulos remark that scientists are obliged to undertake the task of rebutting the "erroneous and harmful antitheses" themselves; I conclude that they should not attempt it alone, and that it would be prudent to read the literature of theologians and moral philosophers, and even to seek their collaboration.

RICHARD C. HARDWICK
17 Duke Street, Leamington Spa,
Warwickshire, CV32 4TR, UK

1. Gödel, K. *On Formally Undecidable Propositions* (Basic Books, New York, 1962).

SIR—Theoharis and Psimopoulos's attack on the philosophy of science and, though it was not named, the sociology of science, was a mediocre affair.

To single out one failing among many, this article all but ignores the debates about the nature of truth that have gone hand in hand with the development of science throughout history. The almost completely unanalysed idea of truth that crops up like a rash all over this article would have been alien to nearly all the great scientists of the past. In fact, in all but one case the word "truth" could apparently be replaced throughout the article (except where it appears in inverted commas) by the word "tradition" without changing the sense at all. The exception is "objective truth", but since in the

authors' terms that is merely a tautology, the same point still holds.

The fact that there is no proper analysis of any of the terms of the argument serves to underline the poverty of the authors' case. They create an extraordinary straw-man opponent, compounded of Popper, Kuhn, the British Broadcasting Corporation and the pre-socratics to name but a few, and ascribe to it views which none of the sources named would necessarily hold. Then, rather than join in argument with this construction, they just inform us that the "antitheses" in question have been adequately "exposed" in a "debunking" elsewhere.

Finally, the sort of attack on philosophy and the human sciences as they impinge upon the natural sciences that this article represents is worrying. The argument that an entire body of knowledge should be suppressed because it interferes with the funding of "proper" science is one that leads to intellectual stalinism. The idea that science should be immune from serious criticism, or even serious analysis, is one that must be seen for what it is: dangerous absolutism.

OLIVER MORTON

1 Grimsby Road,
Louth, Lincolnshire, UK

SIR—Aside from philosophical arguments, it is difficult even on empirical grounds to support the thesis of Theoharis and Psimopoulos that the decline of British science is the result of the denial of objectivity in scientific research. As two of the discredited four "irrationalist" thinkers are Americans, one would have thought that the malaise affecting British science would have spread to the United States. Yet on objective grounds (such as research funding for basic research) this is plainly not true.

There are, however, more serious grounds for rejecting Theoharis and Psimopoulos's thesis. The first is that to accept it may well be in itself an unscientific act. Whether one agrees with their conclusions or not, they have failed to produce any rational argument as to why we should reject the arguments of the anti-objectivists. Their own argument itself rests on a subjective assertion of three stated antitheses, and it is certainly not good science to reject a theory because one doesn't like the conclusions.

And what of the conclusions themselves? In contrast to the authors, it can be argued that "by denying truth and reality" science in fact offers the world one of the great philosophical realizations of this century: that we cannot logically make any assertion with complete confidence. More, we cannot afford to. Truth and tolerance are tightly linked concepts these days, as (largely as a result of the ideas of thinkers such as Popper, Lakatos, Kuhn and Feyerabend) the most basic truth is that there can be no objective truth. And once one accepts that,

it is essential that we tolerate, and even respect, the ideas of others.

The problem with science is really that it is misunderstood. We learn from an early age that science deals in concrete (yes/no) answers, and the consequences of this can be dire in the face of the complex problems of today. We come to expect concrete achievements and in the end believe that all of the world's problems will be solved by the relentless application of the so-called scientific method. But they will not. Perhaps the solution is to give rather more emphasis to the thoughts of the anti-objectivists, rather than suppressing them on anxious subjective grounds.

GREG GIBSON

Biozentrum der Universität Basel,
Klingelbergstrasse 70,
CH-4056 Basel, Switzerland

SIR—Theoharis and Psimopoulos may be more correct than the philosophers they criticize, but to imply that science deals with truth 100 per cent of the time will not restore it to its previous status of credibility. Science works very well, but in spite of error. It does, however, produce firm and lasting contributions to knowledge. Does anyone seriously say the defeat of smallpox was chance? Or that the theories and observations of the Manhattan project will ever be proved erroneous? Although theory may sometimes bias observation, the aim of science is to reduce bias to negligible levels.

So what can we say to the sour puzzlement of philosophers, who cannot understand why the the principles of science work? We reply that the principles of science are natural, essential to everyday life. Consider children, who are built to observe, experiment, learn and apply. Their progress is real enough, even though they apply the principles haphazardly. Adults also apply the principles haphazardly, but the fact that the human race continues to exist shows that the principles work surprisingly well. Science uses the same principles, but with a rigour and uniformity rare in history. It is hardly surprising, therefore, that science works and that it works the better for those who use the principles efficiently. But it is difficult to sell this principle to nonscientists for funding purposes.

Why do scientists think very little about the philosophy of science? Because the activity of science is basically natural. Indeed, the success of science may be merely another application of the Anthropoc Principle. Whether it will continue to be so powerful with phenomena far removed from normal human experience is another question.

N.E. WHITEHEAD

International Laboratory of Marine
Radioactivity,
Musée Oceanographique,
Monaco, MC98000

Nature of the 650-km seismic discontinuity: implications for mantle dynamics and differentiation

A. E. Ringwood & T. Irifune

Research School of Earth Sciences, Australian National University, Canberra, A.C.T. 2601, Australia

Chemical buoyancy relationships during subduction of young, thin oceanic plates cause them to be trapped in a gravitationally stable layer between 600 and 700 km that partly isolates the convective systems of the upper and lower mantle. But when mature, thick plates are subducted, their upper, cool differentiated layers may break through this barrier and become entrained in the convective circulation of the lower mantle. The resultant hybrid convection model provides a promising explanation of the properties of the seismic discontinuity at 650 km depth and the geochemical evolution of the mantle.

THE nature of the major seismic discontinuity occurring near a depth of 650 km has been widely discussed in recent years. According to one view, the discontinuity is caused by an isochemical phase transformation of Mg_2SiO_4 spinel to MgSiO_3 perovskite plus $(\text{Mg},\text{Fe})\text{O}$ magnesiowüstite, occurring within a mantle of approximately uniform composition. This interpretation permits whole-mantle convection to occur freely between the upper and lower mantles. An alternative hypothesis maintains that the discontinuity is associated with a change in chemical composition from an overlying upper mantle dominated by Mg_2SiO_4 olivine and spinel minerals to a relatively silica-rich lower mantle composed essentially of MgSiO_3 perovskite. This interpretation implies that the convective systems of the upper and lower mantles are essentially independent and separated by a thermal boundary layer at ~ 650 km depth.

The rationale for the chemically zoned mantle model has been based upon the hypothesis that the Mg/Si ratio of the bulk mantle is similar to that in C1 chondrites^{1,2}. Extensive melting and differentiation of an essentially chondritic mantle is believed to have occurred during, or soon after, accretion of the Earth, and was dominated by the fractionation of majorite garnet and/or MgSiO_3 perovskite, leading to a lower mantle of metasilicate stoichiometry and an olivine-rich upper mantle predominantly of pyrolite composition^{3,4} (Table 1). However, recent high-pressure, high-temperature experiments by Kato *et al.*^{5,6} have shown that the proposed fractionation processes would have caused the composition of the upper mantle to be quite different from that which is observed. Although a few workers⁷ have claimed that a chemically differentiated mantle is marginally more consistent with its density and velocity distributions, most workers who have examined the issue have concluded that the present resolution of the physical properties of the lower mantle is inadequate to discriminate between the differentiated-mantle and uniform-mantle models (see ref. 8 for a review of this topic).

On the other hand, it has been strongly argued⁹ that the discontinuity at 650 km is too sharp to be caused solely by the isochemical transition of $(\text{Mg},\text{Fe})_2\text{SiO}_4$ spinel to MgSiO_3 perovskite plus $(\text{Mg},\text{Fe})\text{O}$ magnesiowüstite occurring in a chemically uniform mantle of pyrolite composition. It therefore seems that neither of these end-member models (chemically uniform mantle; grossly differentiated mantle) is capable of fully explaining the observed and inferred physical and chemical properties of the mantle. We describe here a hybrid model, which attempts to reconcile these problems and provides a new interpretation of the dynamical behaviour of the mantle. The basic high- P , T experimental data on which this model now rests were obtained quite recently and have been published elsewhere^{10,11}.

When averaged on a large scale (>100 km), the major-element chemical composition of the mantle is considered to be approximately uniform throughout and is represented by the pyrolite model composition⁸ (Table 1). The principal site of medium-scale (5–100 km) differentiation of the mantle occurs at mid-oceanic ridges, where pyrolite partially melts to form a basaltic crust underlain by residual harzburgite. A second region where analogous differentiation occurs is beneath island arcs and back-arc spreading centres. Figure 1 shows an idealized section of the structure of the resultant plates of oceanic lithosphere. Ultimately, the oceanic plates are subducted into the mantle at oceanic trenches. When the differentiated lithosphere encounters the discontinuity at 650 km, a complex interaction ensues that modifies the structure of the discontinuity and may regulate the convective regime of the entire mantle.

Density relationships

As differentiated oceanic lithosphere plates descend into the mantle, they undergo a complex series of successive phase transformations with increasing depth. The sequences and characteristics of these phase transformations are considerably influenced by the differing chemical compositions of the basalt, harzburgite and 'depleted' pyrolite layers (Table 1). This topic has been reviewed in detail by Ringwood^{8,12,13}. The phase transformations exert a strong influence on the density distribution within the subducted lithosphere, and between subducted lithosphere and surrounding mantle. Gravitational body forces arising from these density distributions are expected to play an important role in driving and regulating the subduction process.

We have determined experimentally the phase relationships displayed by the basalt, harzburgite and pyrolite compositions

Table 1 Chemical compositions

	Pyrolite	Harzburgite (depleted)	Basalt (MORB)
SiO_2	44.6	43.6	49.7
Al_2O_3	4.3	0.6	16.4
CaO	3.5	0.5	13.1
MgO	38.1	46.5	10.1
FeO^*	9.1	8.8	8.0
Na_2O	0.4	—	2.0
TiO_2	—	—	0.7
100 $\text{MgO}/$ ($\text{MgO} + \text{FeO}^*$)	88.2	90.5	69.2

Data from refs 8, 11. $\text{FeO}^* = \text{FeO (total)} + \text{NiO} + \text{MnO} + \text{CrO}$.

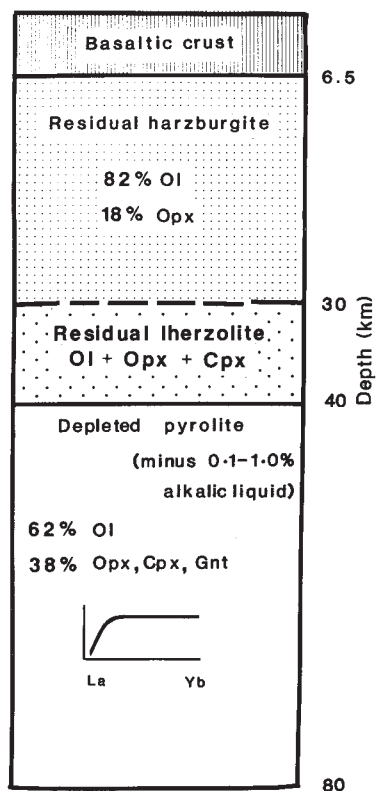


Fig. 1 Idealized structure of oceanic lithosphere showing chemical and petrological zoning developed during partial melting and differentiation at mid-oceanic spreading centres. An analogous structure may form beneath back-arc spreading centres, and the model described here is also applicable to subduction of the highly refractory harzburgite lithosphere developed in island-arc environments. Ol, olivine; Opx, orthopyroxene; Cpx, clinopyroxene; Gnt, garnet.

of Table 1 to a depth of ~ 720 km in the mantle using an MA-8 high pressure apparatus operating up to 26 GPa and 1,600 °C. Densities of the various phase assemblages, reduced to ambient P, T conditions, have been calculated as a function of depth. We have then corrected these 'zero-pressure' densities for compressibility and thermal expansion along a reasonable mantle geotherm¹⁴ using procedures described in refs 11 and 15.

Figure 2a shows the density difference between pyrolite and a primitive mid-ocean-ridge basalt (MORB) along the chosen geotherm. It is seen that the former basaltic crust is denser by $0.1\text{--}0.2\text{ g cm}^{-3}$ throughout most of the depth interval. However, there is a sharp reversal between about 650 and 730 km, where basaltic crust is $\sim 0.08\text{ g cm}^{-3}$ less dense than pyrolite. This confirms a prediction by Anderson¹⁶ and is caused by the fact that the transformation of garnet to perovskites in the basaltic composition occurs over a much broader pressure interval and is not completed until a substantially higher pressure than the equivalent transformation in the pyrolite composition. This is primarily due to the much higher content of Al_2O_3 in basalt than in pyrolite, which extends the stability field of garnet to greater depths. As garnet is $\sim 10\%$ less dense than perovskite, its survival in the basaltic composition in a pressure regime where all of the garnet in pyrolite has transformed to perovskite causes the basaltic composition to be less dense than pyrolite in this same depth interval.

Figure 2b compares the density of harzburgite with that of pyrolite along the same geotherm: harzburgite is $\sim 0.05\text{--}$

0.08 g cm^{-3} less dense than pyrolite to a depth of ~ 500 km. This is due to the smaller iron content of harzburgite, combined with its lower Al_2O_3 content that is responsible, in turn, for a smaller amount of dense garnet. Harzburgite is also $0.05\text{--}0.08\text{ g cm}^{-3}$ less dense than pyrolite below 700 km. The deficit is caused by: (1) the lower FeO content of harzburgite; (2) the larger quantities of CaO and Al_2O_3 in pyrolite, which lead to the formation of ultra-dense CaSiO_3 and aluminous phases; and (3) the lower $\text{MSiO}_3/\text{M}_2\text{SiO}_4$ ratio in pyrolite¹². In the region between 570 and 640 km, where both compositions crystallize mainly to spinel + garnet assemblages, the density of harzburgite is very close to that of pyrolite. In sharp contrast to these regions, in the layer between 650 and 690 km, harzburgite is significantly denser than pyrolite: this is because pyrolite contains more alumina than harzburgite does, which causes garnet to remain stable to greater depths in the former composition^{11,17,18}.

Figure 2c shows the density difference between basalt and harzburgite along the mantle geotherm. Basalt is $0.1\text{--}0.3\text{ g cm}^{-3}$ denser than harzburgite except between 650 and 720 km, where harzburgite is $\sim 0.15\text{ g cm}^{-3}$ denser than basalt. This reversal is a consequence of the stability of garnet to higher pressures in the basalt than in the harzburgite composition, and is analogous to the situation with pyrolite and basaltic crust.

Implications for mantle layering

Figure 2 shows that former harzburgite is gravitationally stable within a pyrolite mantle in the depth interval 650–700 km, whereas former basaltic crust is likewise gravitationally stable in a layer immediately above 650 km. If mantle dynamical processes exist that are capable of transporting large volumes of harzburgite and basalt to a depth of ~ 650 km under conditions where these lithologies are approximately isothermal with surrounding pyrolite mantle, it is possible that they would spread laterally to create stable, globally stratified layers¹⁹. This result could be achieved by several processes related to subduction. For example, if large bodies of harzburgite were subducted into the lower mantle and ultimately reached thermal equilibrium, they would become buoyant (Fig. 2). Such bodies could rise diapirically upwards, becoming gravitationally trapped when they encountered the 640–700-km depth interval. More generally, the return flow in processes of whole-mantle convection would necessarily cause isothermal domains of segregated former harzburgite and basalt to be transported upwards through the layer at 600–700 km where they may also become gravitationally trapped.

The subduction of young (<20 Myr) and thin (<40 km) slabs may provide the major continuing source of harzburgite and basalt for gravitational trapping at these depths. These slabs possess comparatively low thermal inertia and may also be subducted more slowly than older, thicker slabs. They are heated by thermal conduction during descent and their internal temperatures may approach equilibrium with surrounding mantle by a depth of 650 km. Accordingly, they become aiseismic and ductile at shallow depths^{20,21}. Another mechanism for subduction of extremely young (<5 Myr) oceanic lithosphere can occur where there is a variation in age along the trench. Subduction is driven primarily by the gravitational body forces exerted on older lithosphere, which may in turn entrain adjacent younger lithosphere and cause it to be subducted. This process seems to have occurred during subduction of the Farallon plate beneath the coast of NW America.

Figure 2 illustrates the buoyancy relations of thermally equilibrated components of subducted oceanic lithosphere. As shown by Lister and Kerr¹⁹, the former harzburgite and basaltic crust may tend to respond to this buoyancy regime by spreading out to form a gravitationally stable layer separating the upper and lower mantle. The thickness of this layer is likely to grow with time, until it occupies the region of maximum (local) gravitational stability between depths of ~ 600 and 700 km (Figs 2, 3 and 4).

The buoyancy relationships of the former basaltic component of the subducted lithosphere are also interesting in this respect. Former basalt is about 0.15 g cm^{-3} less dense than harzburgite in the perovskite facies, immediately below the discontinuity at 650 km (Fig. 2c). Accordingly, blobs of former basaltic oceanic crust may become trapped on top of the discontinuity^{10,16} and spread out to form lenses (Fig. 5). Seismic velocities increase discontinuously at the interfaces between the bases of these lenses and the underlying lower mantle because of the change in chemical composition. This structure may well explain the evidence that the 650-km discontinuity is sharp enough to reflect seismic waves in some but not all localities^{9,22,23}.

The gravitationally stable layer of former oceanic lithosphere that is trapped between depths of 600 and 700 km (Fig. 5) may provide a considerable degree of isolation between the convective systems of the upper and lower mantle. Once this structure is established by any of the mechanisms discussed above, it provides a barrier to deeper penetration by young, thin slabs. However, for reasons developed below, convective isolation of the two regions is not absolute, because the barrier can still be penetrated by old, thick slabs.

Interactions at the 650-km discontinuity

Most of the total volume of subducted oceanic lithosphere is in the form of mature plates, typically 60–100 km thick and 50–100 Myr old (at the point of subduction)^{20,24}. Mature oceanic lithosphere plates of this type (Fig. 1) possess much greater thermal inertia than the younger, thinner plates discussed in the previous section. In consequence, the interiors of mature plates may be several hundred degrees cooler than surrounding mantle when they encounter the 650-km discontinuity^{25,26}. Because of their lower temperatures, mature plates may display seismic activity to depths of 650 km. Besides exhibiting the capacity for brittle failure in their interiors, the mean viscosities of these plates are likely to be several orders of magnitude higher than

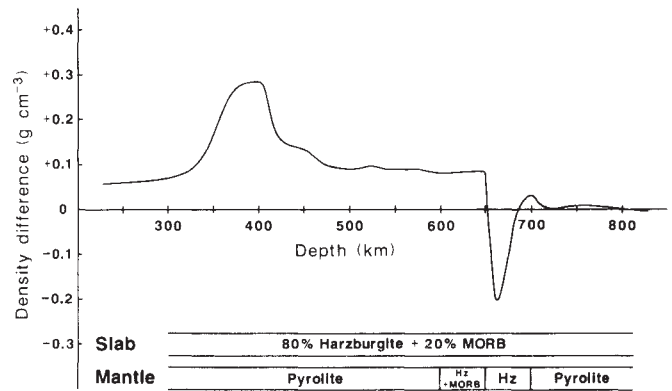


Fig. 3 Density differences between subducted slab and surrounding mantle. The slab is assumed to consist of 20% basalt and 80% harzburgite and to be cooler than surrounding mantle by 800 °C at 400 km and by 400 °C at 650 km, attaining thermal equilibrium with surrounding mantle at ~900 km. The surrounding mantle follows the geotherm of Brown and Shankland¹⁴ and consists of pyrolite above 600 and below 700 km. Between these depths the mantle consists of a pre-existing layer of basalt and harzburgite as shown in Fig. 4 and discussed further in text.

surrounding mantle²⁷. Accordingly, they may behave quite differently from the younger, slower aseismic plates when they encounter the 650-km discontinuity.

The buoyancy relationships between a mature plate and surrounding mantle are shown in Fig. 3, which is based upon data in Fig. 2, and corrected for the effects of lower temperatures in the slab as described in refs 8 and 11. It is assumed that the interior of the plate is ~800 °C cooler than surrounding mantle at a depth of 400 km, ~400 °C cooler at 650 km, and attains thermal equilibrium with surrounding mantle at ~900 km, as discussed in relevant thermal models^{11,25}. The surrounding upper and lower mantle is assumed to be composed of pyrolite except for the layer between 600 and 700 km, which is assumed to consist of mixed former harzburgite and basalt (600–650 km) and harzburgite alone (650–700 km) as discussed in the previous section, and depicted in Fig. 4.

Figure 1 shows the structure of the descending mature plate. It consists of an upper differentiated zone of basalt/harzburgite (0–30 km) and a lower zone, between depths of 40 and 80 km, consisting of slightly depleted pyrolite. The buoyancy relationships in Figs 2 and 3 show that the pyrolite section of the slab is significantly less dense than the former basalt in the layer between 600 and 650 km and is also less dense than the former harzburgite layer between 650 and 700 km. It therefore seems that the differentiated layer of former basalt and harzburgite between 600 and 700 km may provide a barrier to penetration by the depleted pyrolite section of the slab into the lower mantle. We thus suggest that this part of the slab is resorbed into the convective circulation of the upper mantle, as illustrated in Fig. 4. Additional factors that could enhance resorption of the lower oceanic lithosphere within the upper mantle are discussed in refs 8, 12 and 13.

The upper part of the slab is expected to behave quite differently. It is substantially denser than surrounding mantle to a depth of 650 km, mainly because of the effects of lower temperatures in causing thermal contraction and elevation of phase transformations between 300 and 500 km (ref. 12) (Fig. 4). But the basaltic component of the slab is less dense than harzburgite immediately below 650 km (Fig. 2), and the transformation of M_2SiO_4 spinel to (perovskite + magnesiowüstite) has a negative slope of about $-2 \text{ MPa } ^\circ\text{C}^{-1}$ (ref. 18). For a temperature difference of 400 °C, M_2SiO_4 spinel would remain stable in the slab for ~20 km below the 650 km discontinuity.

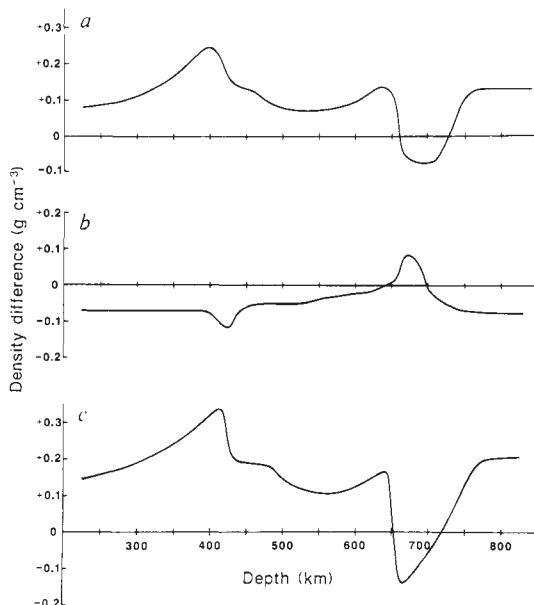


Fig. 2 Density differences between basalt (MORB), harzburgite and pyrolite compositions as a function of depth along a plausible mantle geotherm¹⁴. *a*, MORB minus pyrolite; *b*, harzburgite minus pyrolite; *c*, MORB minus harzburgite. Densities are calculated from experimentally determined phase assemblages in these compositions, corrected for compressibility and thermal expansion along the above geotherm¹¹. These curves are applicable to a subducted slab that has thermally equilibrated with surrounding mantle by ~600 km depth.

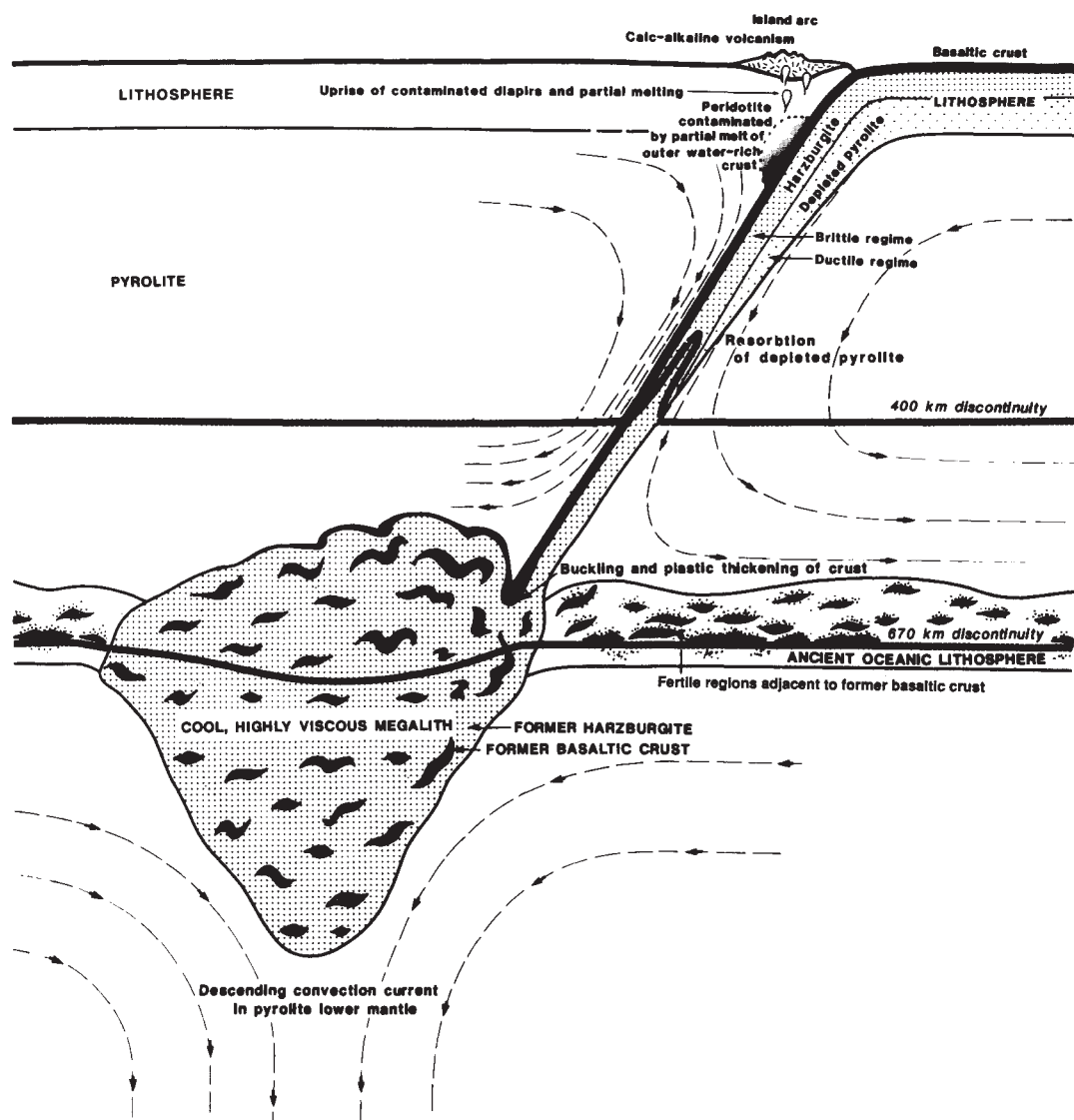


Fig. 4 Model showing subduction of a cool, thick plate of differentiated oceanic lithosphere. Previous subduction episodes involving thin, thermally equilibrated plates have produced a layer of former harzburgite and basalt ('ancient oceanic lithosphere') between 600 and 700 km. The tip of a cool, thick plate experiences buoyant resistance when it penetrates this layer and encounters the discontinuity at 650 km (the 670-km discontinuity in the above figure should read 650-km discontinuity). At that depth the former oceanic crust and harzburgite layers may plastically thicken and buckle to form a large melange (megalith) situated mainly below the seismic discontinuity. The megalith is a transient feature and ultimately becomes entrained in the convective regime of the lower mantle. The lower layer of ductile depleted pyrolite initially at the base of the descending plate of sub-oceanic lithosphere becomes resorbed into the upper mantle by convective circulation owing to its inability to penetrate the harzburgite-basalt layer at 600–700 km because of the buoyancy relationships depicted in Figs 2 and 3.

Throughout this depth interval the slab is up to 0.23 g cm^{-3} less dense than surrounding mantle (Fig. 3). The buoyant stress on the tip of the slab as it penetrates the discontinuity at 650 km is up to 500 bar (0.05 GPa).

Ringwood^{8,12,13} has suggested that under these conditions the tip of the slab may experience buckling and plastic thickening. Further penetration of the slab would be impeded by the high viscosity of the buckled and thickened leading edge, and the intensity of deformation would be enhanced. In consequence, the slab may pile up to form a huge melange or 'megalith' of former basalt and harzburgite^{8,12,13}.

Continuation of this process over 10^8 years would yield a megalith with cross-sectional area $\sim 2 \times 10^5 \text{ km}^2$. The shape of the megalith is likely to be controlled by several factors, particularly its mean temperature and the velocity of migration of the oceanic trench relative to the megalith. A wide range of

geometries is permitted. The particular example illustrated in Fig. 4 corresponds to a situation where the mean temperature of the megalith is $\sim 300^\circ \text{C}$ cooler than the surrounding mantle. Accordingly, it is essentially neutrally buoyant and in isostatic equilibrium^{8,13} (Fig. 4). If the megalith were substantially cooler than the example shown in Fig. 4, it would sink into the lower mantle because of its higher density, and accordingly its vertical dimension would be much greater than its width. On the other hand, a megalith formed by subduction beneath a rapidly migrating trench could be horizontally elongated. Megaliths may be responsible for seismic anomalies observed below the intersections of some slabs with the 650-km discontinuity^{26,28}. The presence of megaliths combined with the complex layered structure illustrated in Fig. 4 is also consistent with the conclusion by Dziewonski and Woodhouse²⁹ that the degree of seismic heterogeneity in the mantle at depths of $\sim 650 \text{ km}$

is greater than at shallower (<500 km) and deeper (>800 km) depths.

Convective regime in the mantle

The region of the megalith below 650 km provides a 'cold finger' or heat sink in the lower mantle. This may be coupled to a sinking convection current in the lower mantle that would erode and entrain the outer shell of the megalith. Ultimately, it seems likely that most of the megalith would become mixed into the lower mantle by convection⁸.

Earlier, it was argued that young plates which approach temperature equilibrium with the mantle at ~650 km may become trapped to form a gravitationally stable chemical and thermal boundary layer separating the upper and lower mantle (Fig. 4). On the other hand, mature plates may buckle and thicken to form megaliths when they encounter this layer. The megaliths puncture the boundary layer and ultimately enter the lower mantle. Mass-balance considerations require that an equivalent volume of pyrolite from the lower mantle would then ascend as plumes through the boundary layer and become mixed into the upper mantle. According to the model, a dynamic equilibrium is established between young subduction processes that establish and maintain a boundary layer of former oceanic lithosphere at ~650 km, and mature subduction processes that cause the formation of megaliths and break through this layer. This process causes substantial exchange of material and mixing between the upper and lower mantle. On the other hand, where the boundary layer of former oceanic lithosphere at ~650 km remains intact, it serves to isolate the upper and lower mantles so that convection occurs separately within these regions.

The Archaean convective regime

Mantle temperatures are believed to have been on average 200–300 °C higher in the Archaean than at present^{30,31}. McKenzie and Weiss³⁰ and Bickle³¹ have argued from thermal considerations that oceanic plate production was much faster in the Archaean than now, that the mean ages of subducted plates were substantially less and that their mean thicknesses were substantially smaller (for instance, 25–40 km) than present oceanic plates. Moreover, the magmatic regimes at Archaean spreading centres are likely to have differed considerably from those today. Higher mantle temperatures would be expected to produce more advanced degrees of partial melting, leading to copious generation of komatiitic magmas³². The refractory residuum remaining after extraction of these magmas would consist primarily of magnesian dunites (Fo₉₄) (ref. 33). The komatiitic magmas are likely to have fractionated nearer the crust-mantle interface by extensive crystallization of olivine, leading to the development of an oceanic crust composed mainly of tholeiitic basalts. Thus the Archaean oceanic lithosphere may have consisted of successive layers of basalt, dunite cumulates and dunitic refractory residua (~Fo₉₄).

This combination of oceanic lithosphere structure and thickness would lead to highly efficient gravitational trapping of subducted plates between 600 and 700 km (Figs 2 and 3). The gravitational stability of Archaean subducted magnesian dunite between 650 and 700 km would have been substantially higher than that of modern harzburgite. Former dunitic (Fo₉₄) oceanic lithosphere would have been 0.12 g cm⁻³ less dense than pyrolite below 700 km and is unlikely to have been subducted into the lower mantle in large volumes. Under these conditions we expect that the tendency for stable stratification of the mantle between 600 and 700 km would have been more strongly developed, with the region at 600–650 km consisting of former basaltic oceanic crust and the region at 650–700 km composed of former magnesian dunites derived from subducted oceanic lithosphere. We suggest that this structure served as a chemical boundary layer isolating the separately convecting upper-mantle and lower-mantle systems throughout most of the Archaean. During this period a large proportion of the present continental crust was

Table 2 Compositions of liquid and residual garnet produced by partial melting of MORB oceanic crust (Table 1) at 20 GPa and ~2,200 °C

	Residual garnet	Liquid
SiO ₂	45.5	54.2
TiO ₂	0.2	1.9
Al ₂ O ₃	20.2	8.4
FeO	0.8*	2.0*
MgO	25.0	13.7
CaO	7.4	16.8
Na ₂ O	0.3	2.6

Data from ref. 6.

* Low values caused by loss of iron to container during experiment.

formed by differentiation and fractionation solely from the upper mantle reservoir, which was continually rehomogenized by convection^{34,35}. At a much later stage, perhaps in the Proterozoic, further cooling of the mantle allowed the formation of thicker (80–100 km) plates of oceanic lithosphere. Some of these plates were sufficiently dense and massive to penetrate the chemical boundary layer at 600–700 km, leading to the present hybrid convective regime, as discussed earlier.

The higher mantle temperatures in the Archaean would have caused more extensive partial melting of subducted basaltic oceanic crust than occurs now. Moreover, remelting may have extended over a much broader depth interval, extending to 600 km. Table 2 shows that the magma produced by partial melting of oceanic crust under these conditions has a higher CaO/Al₂O₃ ratio (2.0) and that the refractory residuum is pyrope garnet with a low CaO/Al₂O₃ ratio (0.37). These melts would have hybridized adjacent regions of the upper mantle, which would have later become homogenized by convection³⁵. Continuation of this process during the Archaean may account for the strong evidence³⁶ that the upper mantle possesses a CaO/Al₂O₃ ratio (0.92) that is ~12% higher than the chondritic value (0.82). The residual, refractory garnet remaining after partial melting would have become trapped in the chemical boundary layer between 600 and 650 km, as previously discussed. Hybridization of adjacent bodies of refractory dunite by these slab-derived magmas possessing high CaO/Al₂O₃ ratios could have produced source regions that later generated Barber-ton-type komatiitic magmas.

Petrogenesis of MORBs

MORBs appear to have formed by partial melting of a pyrolite source region that has experienced episodic extraction of small amounts of alkalic magmas, highly enriched in incompatible elements, over a period exceeding 10⁹ yr (ref. 37). Figures 1 and 4 give a possible explanation of this phenomenon. During the formation and differentiation of oceanic lithosphere, the lowermost portion is believed to have experienced the loss of only a very small proportion of alkalic magma, causing selective depletion of highly incompatible elements. This lower lithospheric layer of slightly depleted pyrolite remains 'fertile' in its capacity to produce tholeiitic magmas when subjected to more extensive degrees of partial melting. However, during subduction, this layer is unable to penetrate the boundary layer of former oceanic lithosphere at ~650 km because of its relative buoyancy, as discussed previously (Figs 2 and 3). Accordingly, it is trapped and recycled into the upper mantle. This process, operating over a timescale of thousands of millions of years, has the capacity to produce the depletion patterns of incompatible elements displayed by MORB source regions. As megaliths ultimately derived from differentiated upper mantle pyrolite are transferred to the lower mantle, replenishment of the MORB source region is achieved by upwelling of plumes of geochemically primitive pyrolite from the lower mantle through the chemical boundary layer at 650 km. Convection within the upper mantle must be

highly efficient, to provide the effective mixing that seems to be required to explain the degree of chemical and isotopic homogeneity displayed by the MORB source region.

Petrogenesis of intraplate volcanic association

Basalts belonging to the intraplate basaltic association occur widely, both in oceanic and continental regions, and include tholeiites, alkali basalts, basanites, nephelinites and kimberlites. The chemical and isotopic signatures of intraplate basaltic rocks imply that their source regions have experienced a much more complex geochemical evolution than the MORB source region. Hofmann and White^{38,39} and Chase⁴⁰, among others, have argued that many of the chemical and isotopic characteristics of these magmas were derived ultimately from former basaltic oceanic crust that had been subducted and stored in the mantle for more than 10^9 yr to form discrete source regions. The present model suggests how this might be accomplished.

Subducted oceanic lithosphere probably contains significant amounts of serpentinite (providing a source of water)¹² and sometimes small amounts of sediment⁴¹. At depths of ~200–600 km, and in the presence of H₂O (and CO₂), bodies of subducted oceanic crust (primarily garnetite–stishovite assemblages) may experience small degrees of partial melting at temperatures well below the solidi of neighbouring harzburgite and pyroxenite⁸. Highly fractionated liquids enriched in incompatible elements would react with, and hybridize, adjacent former harzburgite, thereby transferring their isotopic and trace-element signatures and rendering the former harzburgite 'fertile' in that it would be able to produce alkaline basaltic magmas if subjected to further episodes involving small degrees of partial melting. A significant proportion of this fertile harzburgite may have subsequently become trapped in the boundary layer between 600 and 700 km.

This layer of former oceanic lithosphere also functions as a thermal boundary layer, and over a long timescale (>1 Gyr) is heated from below by the convecting lower mantle. In hotter regions overlying rising lower-mantle convection currents, the upper region of the 'fertilized' boundary layer may become buoyant and ascend into the upper mantle as plumes. The larger plumes would form major 'hotspots' and may be responsible for well-known features such as the Hawaiian intraplate volcanic chain. Smaller plumes would become incorporated in the oceanic and continental lithosphere, forming localized 'fertile' regions that could provide widely dispersed local sources of alkaline magmas when subjected to later melting processes.

Differentiation of crust and mantle

Studies^{42,43} based on trace element and isotopic mass balances provide strong evidence that the continental crust was derived by differentiation from a mantle reservoir that exceeded in volume the upper mantle to a depth of 650 km, but which was much smaller than the volume of the entire mantle. This is difficult to explain, either in terms of simplistic models of whole-mantle convection, or in terms of models in which the convective circulations of upper and lower mantles have remained isolated throughout geological time. The present hybrid model seems to be more compatible with the mass balances required by geochemical data.

Recent studies of U–Th–Pb isotopic systematics in the crust, upper mantle and lower mantle by O'Nions and colleagues^{43,44} throw new light on the geochemical relationships between crust and mantle reservoirs. These authors have concluded that highly incompatible elements such as U, Th and Pb are efficiently extracted and transferred from the upper mantle to the crust on a comparatively short timescale of <600 Myr. Replenishment of these elements occurs by injection of relatively undepleted material from the lower mantle into the upper mantle. The rate of injection of material from the lower mantle is estimated as several per cent of the upper mantle mass per 10^9 years. Efficient convection causes mixing and homogenization of injected undepleted material from the lower mantle with depleted upper-mantle material. O'Nion's model, based upon geochemical arguments, is remarkably consistent with the hybrid model of mantle convection developed in this paper from completely different considerations.

Although our model seems to provide promising explanations of certain puzzling geochemical and geophysical characteristics of the mantle, we recognize that its dynamical basis is speculative, particularly in the assumption made about the dominance of high-amplitude local chemical buoyancy over low-amplitude regional thermal buoyancy. We hope that the ideas presented here will provide a challenge to geophysicists and numerical modellers to examine these issues in greater detail.

We thank Professor J. S. Turner, FRS, Dr G. Houseman and Dr S. Kesson for helpful discussions and constructive criticism. *Note added in press:* Partial melting of basaltic oceanic crust during subduction will eliminate stishovite from the garnetite assemblage above 650 km. This will in turn reduce the density of former oceanic crust overlying the 650 km discontinuity (Fig. 4) by an additional 0.05 g cm^{-3} , thereby further enhancing its gravitational stability in this layer.

Received 18 May; accepted 8 October 1987.

- Liu, L. G. *Geochem. J.* **16**, 287–310 (1982).
- Anderson, D. L. *J. geophys. Res.* **88**, Suppl. B41–B52 1341–1352 (1983).
- Herzberg, C. T. & O'Hara, M. J. *Geophys. Res. Lett.* **12**, 541–544 (1985).
- Anderson, D. L. & Bass, J. D. *Nature* **320**, 321–328 (1986).
- Kato, T., Irifune, T. & Ringwood, A. E. *Geophys. Res. Lett.* **14**, 1546–1549 (1987).
- Kato, T., Ringwood, A. E. & Irifune, T. *Earth planet. Sci. Lett.* (in the press).
- Knittle, E., Jeanloz, R. & Smith, G. *Nature* **319**, 214–216 (1986).
- Ringwood, A. E. in *Proc. Fourth Int. Kimberlite Conf.* Perth (in the press).
- Lees, A. B., Bukowski, M. & Jeanloz, R. *J. geophys. Res.* **88**, 8145–8159 (1983).
- Irifune, T. & Ringwood, A. E. in *High-Pressure Research in Mineral Physics* (eds Manghni, M. & Syono, Y.) 235–246 (Am. Geophys. Un., Washington, 1987).
- Irifune, T. & Ringwood, A. E. *Earth planet. Sci. Lett.* **86** (2) (in the press).
- Ringwood, A. E. *J. Geol.* **90**, 611–643 (1982).
- Ringwood, A. E. *Terra cogn.* **6**, 67–77 (1986).
- Brown, J. M. & Shankland, T. J. *Geophys. J. R. astr. Soc.* **66**, 579–596 (1981).
- Irifune, T. *Phys. Earth planet. Int.* **45**, 324–336 (1987).
- Anderson, D. L. *Science* **223**, 347–355 (1984).
- Liu, L. G. *Earth planet. Sci. Lett.* **36**, 237–245 (1977).
- Ito, E. & Yamada, H. in *High Pressure Research in Geophysics* (eds Akimoto, S. & Manghni, M.) 405–419 (Center for Academic Publishing, Tokyo, 1982).
- Lister, J. R. & Kerr, R. C. *Earth planet. Sci. Lett.* **85**, 241–247 (1987).
- Molnar, P., Freedman, D. & Shih, J. *Geophys. J. R. astr. Soc.* **56**, 41–54 (1979).
- Wortel, R. *Nature* **296**, 553–556 (1982).

- Muirhead, K. J. *geophys. Res.* **90**, 2057–2059 (1985).
- Paulssen, H. *Geophys. Res. Lett.* **12**, 709–712 (1985).
- Vlaar, N. J. & Wortel, M. J. R. *Tectonophysics* **32**, 331–351 (1976).
- Oxburgh, E. R. & Turcotte, D. L. *Bull. geol. Soc. Am.* **81**, 1665–1688 (1970).
- Creager, K. C. & Jordan, T. H. *J. geophys. Res.* **91**, 3573–3590 (1986).
- Carter, N. *Rev. Geophys. Space Phys.* **14**, 301–360 (1976).
- Silver, P. G. & Chan, W. W. *J. geophys. Res.* **91**, 13,787–13,802 (1986).
- Dziewonski, A. M. & Woodhouse, J. H. *Science* **236**, 37–48 (1987).
- McKenzie, D. & Weiss, N. *Geophys. J. R. astr. Soc.* **42**, 131–174 (1975).
- Bickle, M. *Earth planet. Sci. Lett.* **40**, 301–315 (1978).
- Nisbet, E. G. in *Komatiites* (eds Arndt, N. & Nisbet, E.) 501–520 (Allen & Unwin, London, 1982).
- Green, D. H. *Geology* **3**, 15–18 (1975).
- Allègre, C. J. & Turcotte, D. L. *Geophys. Res. Lett.* **12**, 207–210 (1985).
- Kellogg, L. H. & Turcotte, D. L. *Earth planet. Sci. Lett.* **81**, 371–378 (1987).
- Palme, H. & Nickel, K. *Geochim. cosmochim. Acta* **49**, 2123–2132 (1985).
- Gast, P. W. *Geochim. cosmochim. Acta* **32**, 1057–1086 (1968).
- Hofmann, A. W. & White, W. M. *Carnegie Inst. Wash. Yb* **79**, 477–483 (1980).
- Hofmann, A. W. & White, W. M. *Earth planet. Sci. Lett.* **57**, 421–436 (1982).
- Chase, C. G. *Earth planet. Sci. Lett.* **52**, 277–284 (1981).
- Sun, S. S. & McDonough, W. in *Magmatism in Ocean Basins* (eds Saunders, A. D. & Worri, M. J.) (in the press).
- Zindler, A. & Hart, S. A. *Rev. Earth planet. Sci.* **14**, 493–571 (1986).
- Galer, S. J., Goldstein, S. L. & O'Nions, R. K. *Lunar Planet. Sci.* **17**, 247–248 (1986).
- O'Nions, R. K. *J. geol. Soc., Lond.* **144**, 259–274 (1987).

Multiple potassium-channel components are produced by alternative splicing at the *Shaker* locus in *Drosophila*

Thomas L. Schwarz, Bruce L. Tempel, Diane M. Papazian, Yuh Nung Jan & Lily Y. Jan

Howard Hughes Medical Institute and Departments of Physiology and Biochemistry, University of California, San Francisco, California 94143, USA

At least four probable components of potassium channels are encoded at the Drosophila Shaker locus, by a family of alternatively spliced transcripts. Alternative splicing may provide one mechanism of generating the remarkable diversity of potassium channels.

POTASSIUM channels probably form the most diverse group of ion channels, and are essential to the control of the excitability of nerve and muscle. Some potassium channels open in response to a depolarization of the membrane¹, others to a hyperpolarization² or an increase in intracellular Ca^{2+} (ref. 3). Some can also be regulated by the binding of a transmitter⁴ and by intracellular kinases⁵, GTP-binding proteins⁶ or other second messengers⁷. They also differ in their ability to sustain a current or to inactivate, and can be further divided into subtypes according to their kinetic properties and single channel conductances⁸. By their cellular and subcellular distribution, they confer on each neuron its particular properties of excitability and plasticity^{9,10}.

Implications for A channels

The biochemical nature of potassium channels is poorly understood, but the cloning of the *Shaker* locus of *Drosophila melanogaster*¹¹⁻¹⁴ has recently provided a starting point for molecular studies of one type of potassium channel called the A channel. These channels are activated by membrane depolarization and inactivate rapidly^{15,16}. They are frequently encountered in both vertebrates¹⁷⁻¹⁹ and invertebrates^{15,16,20,21}, and influence the rate of firing and duration of action potentials^{20,22,23}, and the amount of transmitter released by terminals²⁴. They have also been implicated in model systems of learning²⁵. *Shaker* mutations can abolish the A current without affecting other currents, including other potassium currents^{20,26,27}. Therefore, the *Shaker* locus was thought to code for a structural component of the A channel. The sequence of a complementary DNA (ShA) from the *Shaker* locus supported this hypothesis¹²; it predicted a protein with six hydrophobic stretches that are likely to be membrane-spanning and an arginine-rich sequence similar to ones that have been found in sodium channels²⁸ and a dihydropyridine-binding protein that is probably a calcium channel²⁹. Also, in the accompanying article we show that RNA transcribed from *Shaker* cDNA can produce A currents when injected into *Xenopus* oocytes³⁰.

As will be shown here, the *Shaker* locus is a complex transcription unit, encoding several different proteins that probably constitute multiple A-channel components. These proteins are encoded by a family of transcripts that arises by alternative splicing; different combinations of exons are selected, as the primary *Shaker* transcript is processed into mature messenger RNA. This is the first example of one channel gene producing several variants of a protein. These proteins appear to contribute to different A-channel subtypes and thereby they offer a biochemical explanation for some of the diversity of this class of potassium channels.

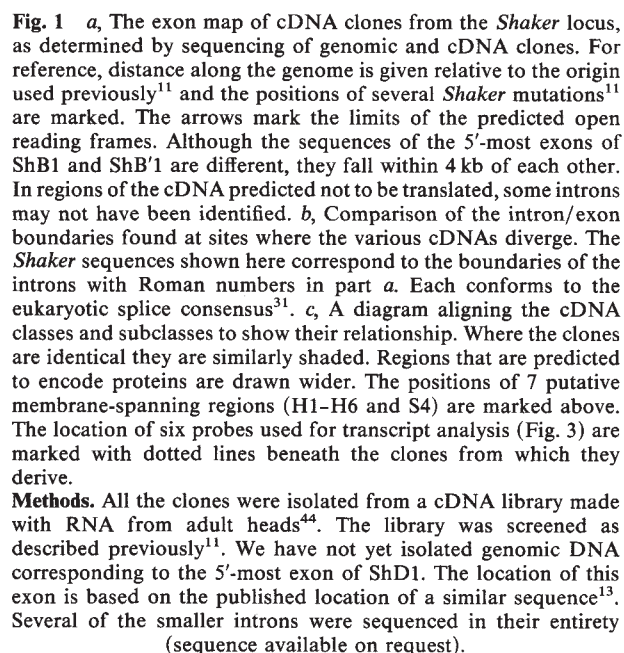
Classes of *Shaker* cDNA clones

We have now sequenced seven cDNA clones that represent five distinct transcripts and predict four different proteins (Figs 1 and 3). We have also addressed the following questions: (1) do the different classes of cDNA arise by alternative splicing, (2) how complex is the pattern of transcription from the *Shaker* locus, (3) are each of the predicted proteins likely to be channel components and how do they differ, and (4) what might be the significance of this protein family?

To describe the different cDNA clones we have adopted the following nomenclature. Clones have been divided into classes on the basis of the proteins they predict (described below) and each class has been given a letter (for example ShA and ShB). Clones that differ only in portions that we predict to be untranslated are further distinguished by a superscript (for example subclasses ShB and ShB'). Finally, each independently obtained clone of a particular class has been numbered (for example ShA1 and ShA2).

In addition to the cDNA clones whose sequence was previously reported (ShA1 and ShA2)¹² we have now sequenced two clones of the class ShB, two of ShC and one of ShD (Fig. 3). The position and size of the exons that encode these cDNA were determined first by hybridizing these cDNA to blots containing cloned genomic DNA, and then by sequencing the corresponding regions of the genome (Fig. 1a). Eight exons that are clustered within 10 kilobases (kb) of the genome are found in all seven cDNA clones. This common region consists of 1,163 nucleotides that constitute roughly the middle third of each of the seven clones. To either side of this region, additional exons from one or more classes of cDNA are spread over a much larger portion of the locus. On the 3' side of the common region, two sets of mutually exclusive exons were observed; one set is shared by classes ShA and ShC, the other by ShB and ShD. Spliced to the common region at its 5' end, three different exons were found; that of ShC, that of ShD, and that shared by ShA and ShB. Still further in the 5' direction, an additional divergence was seen (ShB'1).

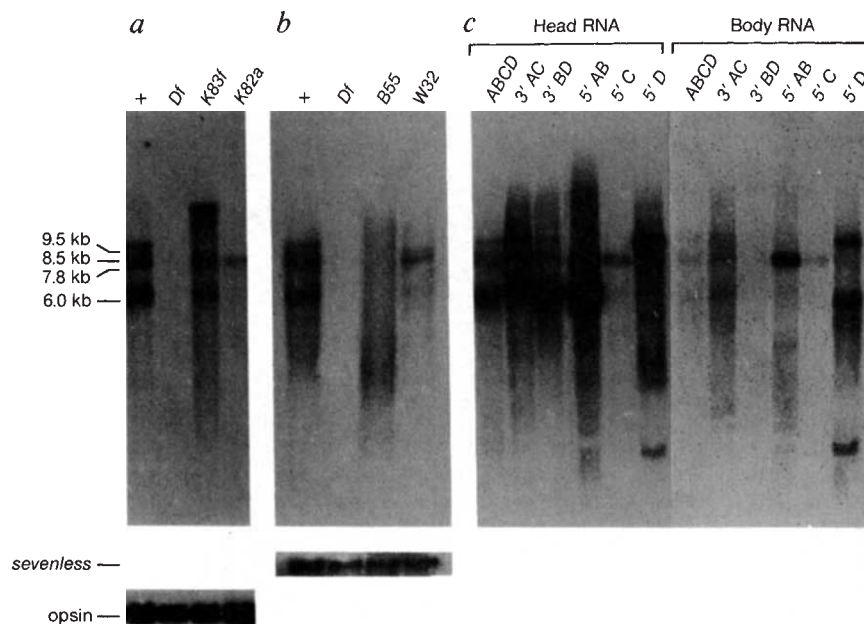
Our data suggest that these many cDNA clones reflect alternative splicing of a large primary transcript rather than precursors or artefactual rearrangements of a single, correctly processed transcript. First, most of the splice junctions are observed in more than one independent cDNA clone. Second, at each of the putative exon junctions the genomic sequence flanking the exon was in accord with the eukaryotic splice-site consensus sequence³¹. In Fig. 1b, the sequences are presented for each of the crucial junctions, that is those where the various classes diverge. Third, the exon map in Fig. 1a is such that no cDNA



We have made a preliminary examination of the tissue distribution of the different *Shaker* transcripts by comparing the transcripts observed in the head with those found in the body (thorax and abdomen) of adult flies (Fig. 2c). Although all transcripts were about 10-fold less abundant in the body, the relative abundance of particular species differed in the two sources. (Compare, for example, the transcripts seen by 3'AC and 3'BD or the two bands seen by 5'AB in Fig. 2c). As will be discussed later, these variants suggest that more than one subtype

Do all these transcripts arise at the *Shaker* locus, or are some due to cross-hybridizing signals from related genes? The existence of numerous *Shaker* mutations allowed us to answer this question and to examine how various mutations altered the transcripts (Fig. 2*a* and *b*). One mutant lacks a large portion of the *Shaker* locus (the region between two breakpoints generated by translocations between the X and Y chromosomes:

Fig. 2 Transcription from the *Shaker* locus. *a*, *b*, Transcripts either from fly heads *a* or whole adult flies *b* are compared with those in *Shaker* mutants. The probe, which contained sequences present in all seven cDNA clones, was a *Pst*I fragment of ShB1 (dotted line labelled ABCD in Fig. 1c). Transcript lengths were determined with a ladder of RNA standards (Bethesda Research Labs). To check that equal amounts of RNA were being compared and that differences were not due to degradation, blots were reprobed with probes to *sevenless*^{47,48} (*a*) or fly opsin (*b*). *c*, A comparison of transcripts seen by probes from different subsets of cDNA. One probe (ABCD) contained sequence in all the cDNA classes we have found. The other probes were made to one of the variant ends: the 3' end shared by classes ShA and ShC (3' AC); the 3' end shared by classes ShB and ShD (3' BD); the 5' end shared by classes ShA and ShB (5' AB); and the 5' end found only in ShC (5' C); the 5' end found only in ShD (5' D). The positions of each probe within the cDNA clone from which it is derived are shown as dotted lines in Fig. 1c. Each probe was also used on blots of RNA from mutants to see if all the signals were authentic *Shaker* transcripts (data not shown). Only the anomalously small band (1.3 kb) seen with probe 5' D was unaltered, even in the deficiency mutant. At present we do not know the origin of this band.



Methods. Flies were frozen in liquid nitrogen and, in *a* and *c*, vortexed and sieved to separate heads from bodies. From each source, the tissue was pulverized while frozen and then extracted with greater than four volumes of a 1:1:2 mixture of phenol: chloroform: 7 M urea, 2% sodium dodecyl sulphate (SDS) and 50 mM Tris, pH 7.5. The aqueous phase was re-extracted six times with 1:1 phenol: chloroform and precipitated with ethanol. Total RNA was re-suspended in 1% SDS, 10 mM HEPES, pH 7.6, and 1 mM EDTA. Polyadenylated RNA was selected on Type II oligo(dT)-cellulose (Collaborative Research). Per lane, 5 µg of head RNA was used or 10 µg of RNA from the other sources. After electrophoresis on formaldehyde gels containing 0.6% agarose, RNA was transferred to Gene Screen nylon membranes (NEN/Dupont) by capillary action and cross-linked to the membrane by UV irradiation⁴⁹. *Shaker* probes were single stranded and made from M13 templates. Control probes were made with hexanucleotide primers on double-stranded templates. Hybridization was in 50% formamide, 5 × SSC, 5% SDS, 5 × Denhardt's at 50 °C overnight; filters were washed for two or more hours at 65 °C in 0.1 × SSC and 0.1% SDS. In *c*, all the lanes are from the same gel and the same probes were used on head and body lanes. Head lanes were exposed for 15 h, body lanes for 5 days. Probe 3' AC was made to a *Bcl*I-*Eco*RI subclone of ShA1, probe 3' BD to a *Pst*I subclone of ShB1, and probe 5' AB to an *Xho*I-*Eco*RI subclone of ShA1. Probes 5' C and 5' D were made by priming ShC1 and ShD1 with oligonucleotides just 5' to the position where their sequences diverge.

of channel are produced by the locus and that these subtypes have different tissue distributions.

Predicted *Shaker* proteins

The different classes of cDNA encode different protein products (Figs 1c and 3). Except for ShD1, all cDNA clones contain at both ends numerous stop codons in all reading frames and a single, large, open reading frame in between. The open reading frame of the ShA class has been discussed previously¹². At the 5' end, ShB1 is identical to ShA and so translation of ShB products probably begins at the same methionine residue proposed for ShA. From this site, the open reading frame of the ShB class continues for 1,968 nucleotides and predicts a protein of 656 amino acids ($M_r = 74,000$). ShC clones are identical to ShA for most of their protein coding region, but predict a different amino terminus. Two closely spaced AUG codons, either or both of which could be translational start sites, are found near the start of the open reading frame in ShC so that a protein of 554 or 556 amino acids may be produced ($M_r = 64,000$). ShD1 is identical to the ShB clones for most of its length, but contains a unique 5' end; the reading frame is open all the way to the 5' end of the clone. Within the sequence that is unique to ShD there are five AUG codons that are potential translational start sites. Alternatively, the start site for ShD may not be present in our clone.

The sequence of another cDNA from the *Shaker* locus has recently been reported¹³. This clone, called F-c, most closely resembles our ShD1, but there are several nucleotides lacking in F-c which cause changes of reading frame. Also, at its 3' end it concludes with a sequence that, by sequencing genomic DNA,

we have determined to be an intron (number 6 in Fig. 3) that has been spliced out of all our cDNA clones. Another recently reported clone¹⁴ predicts another amino-terminal variant which we have not yet encountered. Curiously, this clone terminates midway through the common region, after entering an intron (number 8 in Fig. 3).

The exon map in Fig. 1a and the diagram in Fig. 1c summarize the relationship of the proteins predicted by the four classes of cDNA that we have sequenced. All share a central core. For the carboxy-terminal third, two versions have been found: that in ShA and ShC and that in ShB and ShD. At the amino-terminal end there are three variants that diverge at a single splice site: an eight residue terminus in ShC, a 60 amino-acid sequence in ShA and ShB, and the possibly incomplete version of ShD. No strong similarity of one amino-terminal variant to another was observed. The carboxy-terminal variants, however, were very similar as discussed below.

Closely related channel components

The hypothesis that the *Shaker* locus encodes a structural component of a potassium channel was first advanced on the basis of genetic and physiological studies^{21,26,27}. From the sequence of ShA, two arguments were used in support of this hypothesis¹². Firstly, ShA contained six regions whose length and hydrophobicity indicated that they may be membrane-spanning. Secondly, it contained a region similar to the vertebrate sodium channel, an arginine-rich region called 'S4-like', that has been hypothesized to be a transmembrane helix that is responsible for channel opening in several voltage-dependent channels³³⁻³⁶. These arguments hold equally for the predicted products of ShB, ShC and

ShA 433 FWWAAVVTIMTTVGYGDMTPVGFWGKIVGSLCvvAGVLTIALPVPPIVNSFNFYFHREaDrEEMQSONFNHVTIS
ShB 433 FWWAAVVTIMTTVGYGDMTPVGvWGKIVGSLCaIAGVLTIALPVPPIVNSFNFYFHREtDqEEMQSONFNHVTIS
 FWWAAVVTIMTTVGYGDMTPVG-WGKIVGSLC--AGVLTIALPVPPIVNSFNFYFHRE-D-EEMQSONFNHVTIS

505 CsYLPGaLQGHlKKSSLSSESSDmDLDDGidatTPGLT dHtGR hmvPFL rtQQ sfekqq
 505 CpYLPgtLQGHmKKSSLSSESSDmDLDDG vesTPGLTetHGRsavaPFLGaaqqqqqQpvaslsmsid
 C-YLPG-LQGH-KKSSLSSESSD-MDLDDG---TPGLT-H-GR---PFL-----QQ-----

565 lQLQlqLQqqsQsphgQmTQQQOlqONG lrstnslQL RHINNA
 576 kQLQhPLQhvtQtqlyQQqqQQQQqQNGfkqqqqqtqqQLqqqqshltinasaaaaatsygssglTmRHINNA
 -QLQ--LQ--Q--QQ--QQQQ--QNG--QL-----RHINNA

608 mAVSLETDV
 648 lAVSLETDV
 -AVSLETDV

Even where these two amino acid sequences are most divergent, some similarity is seen. This is due in large part to repeats of the sequence CAX which give rise to an abundance of glutamines in both sequences, but especially in ShB and ShD. Similarly, in their amino-terminal regions ShA and ShB are 41% glutamine from residues 20–54. CAX repeats have been seen in both the translated and untranslated regions of a number of other genes and are termed OPA or M sequences³⁸. Their function is not known.

Significance of alternative splicing

The production of multiple-channel components by alternative splicing is unprecedented, although this mechanism has been encountered frequently in other genes. Alternative splicing can affect untranslated parts of a gene or it can produce variant protein products³⁹. In *Shaker*, both phenomena have been seen and additional splicing alternatives and predicted proteins probably remain to be found. Complex transcription units such as this often involve multiple initiation sites for transcription; different tissues may use different promoters to create distinct primary transcripts that are then distinguished further by alternative splicing⁴⁰. It is tempting to speculate that some of the complexity of *Shaker* reflects the existence of many cell-type specific promoters in the nervous system, as well as specific splicing mechanisms.

We do not yet know why there are multiple *Shaker* proteins, but there are at least two possibilities. First, the *Shaker* products resemble one of the four internally similar domains of the vertebrate sodium channels²⁸ and we have argued that a monomer of a protein such as ShA is probably not a complete potassium channel¹². If, like the vertebrate acetylcholine receptor⁴¹, the complete channel comprises multiple, related subunits, some or all of the four *Shaker* variants could represent these subunits. The second possibility is that these variants may contribute to different potassium-channel subtypes, that is to chan-

nels that differ in their kinetics, regulation and cellular or sub-cellular distribution. Because the relative abundance of the various *Shaker* transcripts differed in the head and body of the fly, it is likely that the *Shaker* proteins do not exist in a fixed stoichiometry in all tissues. Thus some of the *Shaker* products are probably parts of channel subtypes with a differential tissue distribution. Expression in *Xenopus* oocytes of the products of two of these clones supports our hypothesis that they can form A channels whose properties differ³⁰.

Because detailed electrophysiological studies have so far been restricted to a few tissues, little is known of A-channel subtypes in *Drosophila*. *Shaker*-derived A-channels have been studied mostly in muscle^{21,26,27}. Studies have also shown that the electrical properties of larval motor-nerve terminals²⁴ and adult cervical giant fibres²³ are altered in *Shaker* mutants but the channels responsible for this have not been identified. An A-channel subtype has been observed in dissociated embryonic neurons but this channel appears not to be encoded by *Shaker*⁴². Thus, in addition to those that arise by alternative splicing, more members of the potassium-channel family are likely to come from related genes⁴³.

The cloning of *Shaker* has provided a first step in the molecular study of potassium channels. Also, it has yielded the first example of multiple channel components arising from a single gene by alternative splicing of transcripts and suggests one mechanism by which some of the great diversity of potassium channels arises.

We thank K. Lee, L. Ackerman, L. Schulte, M. Doherty and D. Lantz for assistance, Dr P. Salvaterra for the cDNA library and Dr G. Rubin and his colleagues for giving us clones for *sevenless* and *fly opsin* and for advice on RNA extraction. This work was funded by the Howard Hughes Medical Institute and the NIH. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries.

Received 6 October; accepted 18 November 1987.

- Hodgkin, A. L. & Huxley, A. F. *J. Physiol., Lond.* **117**, 500–544 (1952).
- Byerly, L. & Masuda, M. O. *J. Physiol., Lond.* **288**, 263–284 (1979).
- Meech, R. W. & Standen, N. B. *J. Physiol., Lond.* **249**, 211–239 (1975).
- Adams, P. R., Brown, D. A. & Constanti, A. *J. Physiol., Lond.* **330**, 537–572 (1982).
- Grega, D. S., Werz, M. A. & MacDonald, R. L. *Science* **235**, 345–348 (1987).
- Dunlap, K., Holz, G. G. & Rane, S. G. *Trends Neurosci.* **10**, 241–244 (1987).
- Piomelli, D. *et al. Nature* **328**, 38–43 (1987).
- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, Massachusetts, 1984).
- Belardetti, F., Schacher, S. & Siegelbaum, S. A. *J. Physiol., Lond.* **374**, 289–313 (1986).
- Grissmer, S. *J. Physiol., Lond.* **381**, 119–134 (1986).
- Papazian, D. M., Schwarz, T. L., Tempel, B. L., Jan, Y. N. & Jan, L. Y. *Science* **237**, 749–753 (1987).
- Tempel, B. L., Papazian, D. M., Schwarz, T. L., Jan, Y. N. & Jan, L. Y. *Science* **237**, 770–775 (1987).
- Kamb, A., Iverson, L. E. & Tanouye, M. A. *Cell* **50**, 405–413 (1987).
- Baumann, A. *et al. EMBO J.* **6**, 3419–3429 (1987).
- Hagiwara, S. & Sato, N. *J. Physiol., Lond.* **148**, 161–179 (1959).
- Connor, J. A. & Stevens, C. F. *J. Physiol., Lond.* **213**, 21–30 (1971).
- Ito, I. & Maeno, T. *J. Physiol., Lond.* **373**, 115–127 (1986).
- Cassell, J. F., Clark, A. L. & McLachlan, E. M. *J. Physiol., Lond.* **372**, 457–483 (1986).
- Kawa, K. *J. Physiol., Lond.* **385**, 189–205 (1987).
- Connor, J. A. *Fedn Proc.* **37**, 2139–2145 (1978).
- Salkoff, L. *Cold Spring Harb. Symp. quant. Biol.* **48**, 221–232 (1984).
- Connor, J. A. & Stevens, C. F. *J. Physiol., Lond.* **213**, 31–53 (1971).
- Tanouye, M. A., Ferrus, A. & Fujita, S. C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6548–6552 (1981).
- Jan, Y. N., Jan, L. Y. & Dennis, M. J. *Proc. R. Soc. B* **198**, 87–108 (1977).
- Alkon, D. L., Lederhendler, I. & Shoukimas, J. L. *Science* **215**, 693–695 (1982).
- Timpe, L. C. & Jan, L. Y. *J. Neurosci.* **7**, 1307–1317 (1987).
- Wu, C. F. & Haugland, F. J. *J. Neurosci.* **10**, 2626–2640 (1985).
- Noda, M. *et al. Nature* **312**, 121–127 (1984).
- Tanabe, T. *et al. Nature* **328**, 313–318 (1987).
- Timpe, L. C. *et al. Nature* **331**, 143–145 (1988).
- Mount, S. M. *Nucleic Acids Res.* **10**, 459–472 (1982).
- Grenningloh, G. *et al. Nature* **328**, 215–220 (1987).
- Greenblatt, R. E., Blatt, Y. & Montal, M. *FEBS Lett.* **193**, 125–134 (1985).
- Guy, H. R. & Seetharamulu, P. *Proc. natn. Acad. Sci. U.S.A.* **83**, 508–512 (1986).
- Noda, M. *et al. Nature* **320**, 188–192 (1986).
- Catterall, W. A. *Rev. Biochem.* **55**, 953–985 (1986).
- Krebs, E. G. & Beavo, J. A. *Rev. Biochem.* **48**, 923–958 (1979).
- Wharton, K. A., Yedvobnick, B., Finnerty, V. G. & Artavanis-Tsakonas, S. *Cell* **40**, 55–62 (1985).
- Leff, S. E., Rosenfeld, M. G. & Evans, R. M. *Rev. Biochem.* **55**, 1091–1117 (1986).
- Jorgensen, E. M. & Garber, R. L. *Genes Dev.* **1**, 544–555 (1987).
- Noda, M. *et al. Nature* **302**, 528–532 (1983).
- Solc, C. K., Zagotta, W. N. & Aldrich, R. W. *Science* **236**, 1094–1098 (1987).
- Ganetzky, B. & Wu, C. F. *Rev. Genet.* **20**, 13–44 (1986).
- Itoh, N., Salvaterra, P. & Itakura, K. *Drosoph. Inf. Serv.* **61**, 89 (1985).
- Sanger, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Chen, E. Y. & Seeburg, P. H. *DNA* **4**, 165–170 (1985).
- Hafen, E., Basler, K., Edstroem, J. E. & Rubin, G. M. *Science* **236**, 55–63 (1987).
- Banerjee, U., Renfranz, P. J., Pollock, J. A. & Benzer, S. *Cell* **49**, 281–291 (1987).
- Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1995 (1984).
- Sobel, E. & Martinez, H. M. *Nucleic Acids Res.* **14**, 363–374 (1985).

Expression of functional potassium channels from *Shaker* cDNA in *Xenopus* oocytes

Leslie C. Timpe, Thomas L. Schwarz, Bruce L. Tempel, Diane M. Papazian, Yuh Nung Jan & Lily Y. Jan

Howard Hughes Medical Institute and Departments of Physiology and Biochemistry, University of California, San Francisco, California 94143, USA

The Shaker gene of Drosophila melanogaster encodes a potassium-selective ion channel, the 'A' channel, or one of its subunits. A single Shaker messenger RNA species suffices to direct the synthesis of functional A channels in Xenopus oocytes. Physiological characteristics of the A currents induced by two different mRNA species are compared.

VOLTAGE-DEPENDENT potassium channels determine to a large extent the duration of action potentials, the firing patterns of neurons¹, and may be involved in alterations of synaptic efficacy in learning^{2,3}. Voltage clamp studies on *Drosophila* muscle⁸⁻¹¹ indicated that the product of the *Shaker* gene is the 'A'-type potassium channel^{12,13}, or one of its components. Although no potassium channels have yet been purified, the *Shaker* locus has recently been cloned⁴⁻⁷ and the predicted gene

products show sequence similarity with the voltage-sensitive sodium channel and the dihydropyridine receptor of rabbit muscle^{14,15}. These studies, however, have provided only indirect evidence that *Shaker* encodes an A channel, or one of its components, and do not rule out the possibility that one or more polypeptides encoded by other genes are also necessary components of the A channel. At least five different protein products appear to be derived from the *Shaker* gene by alternative

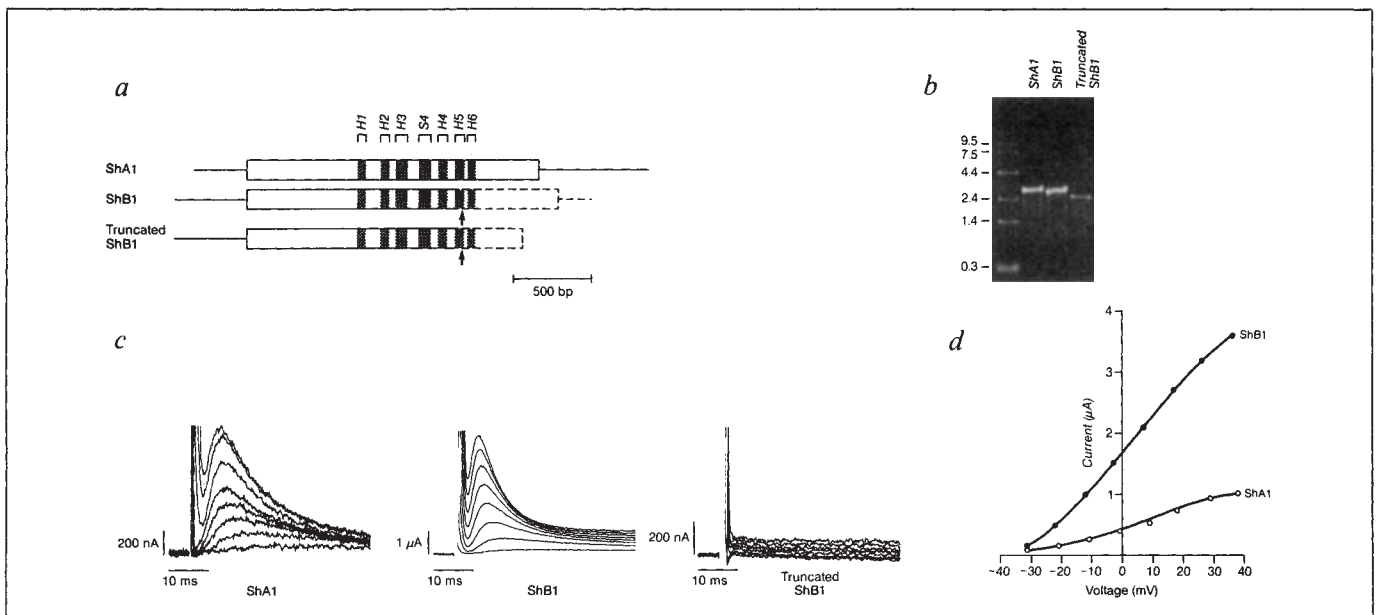


Fig. 1 Expression of transient outward currents in oocytes injected with *Shaker* mRNA. *a*, cDNA templates used in transcription reactions. Boxes indicate open reading frames; thin lines represent untranslated regions^{6,16}. The proposed membrane-spanning hydrophobic sequences and the S4-like sequence are shaded and labelled above. Dashed lines show the portions of the ShB1 cDNA and the truncated ShB1 control whose sequences differ from ShA1. Arrows mark the point of divergence. *b*, Result of *in vitro* transcription of *Shaker* cDNAs. Samples of the mRNA (3–5 µg) were denatured by heating to 65 °C in 50% formamide, then electrophoresed alongside RNA size standards (BRL) on a 0.9% agarose, 7% formaldehyde gel and stained with ethidium bromide. Transcription of the ShA1 and ShB1 templates yielded mRNA of slightly more than 2.4 kilobases (kb), as expected from their sizes (2.9 and 2.7 kb, respectively). The truncated ShB1 template gave a somewhat smaller message, as expected. *c*, Transient outward currents recorded in oocytes at 12 °C. mRNA was taken up in distilled water at 0.1 to 0.2 µg µl⁻¹; ~50 nl was injected per oocyte. The ShB1 and control oocytes received about 5 ng mRNA each, whereas the ShA1 oocytes received either 5 or 10 ng. The currents were large on the second day after injection (the earliest we looked) and persisted for at least five days. Traces show the membrane currents evoked by steps from -80 mV to levels between -30 mV and +40 mV, in ~10 mV increments. A linear leak current, determined from hyperpolarizing steps, has been subtracted. *d*, Peak current-voltage relation. The peaks of the transient outward currents shown in *c* are plotted against the membrane potential. Although a quantitative comparison was not made, we found that oocytes injected with ShA1 mRNA consistently had smaller currents than those which received ShB1 mRNA.

Methods. The ShA1 and ShB1 cDNAs were subcloned into an *Eco*R1 site in the polylinker of a plasmid expression vector (SP72, Promega); these vectors were then linearized by cutting at the *Hind*III site in the polylinker for ShA1 and ShB1, or at a *Pvu*I site 1721 bp into the open reading frame for the control. Run-off transcriptions were carried out with 1–2 µg template DNA, 500 µM NTPs, 500 µM G(5')ppp(5')G (a cap analogue) and 20 units of T7 polymerase in a standard transcription buffer²⁰. After 2 h at 40 °C the template DNA was digested with DNase (RQ1 DNase, Promega), and the protein extracted with 1:1 phenol:chloroform. Nucleic acids were precipitated with ethanol, dried and then taken up in distilled water for injection into *Xenopus laevis* oocytes by a standard method²¹. For culturing oocytes and for all voltage-clamp experiments, except as noted for the tail-current reversal experiments, we used modified Barths saline¹⁷ containing (in mM): 88 NaCl, 1 KCl, 2.4 NaHCO₃, 0.82 MgSO₄, 0.33 Ca(NO₃)₂, 0.41 CaCl₂ and 10 HEPES at pH 7.5. For voltage clamping two 3 M KCl microelectrodes with 1–5 MΩ tip resistances were used; otherwise the techniques were as described previously¹⁰.

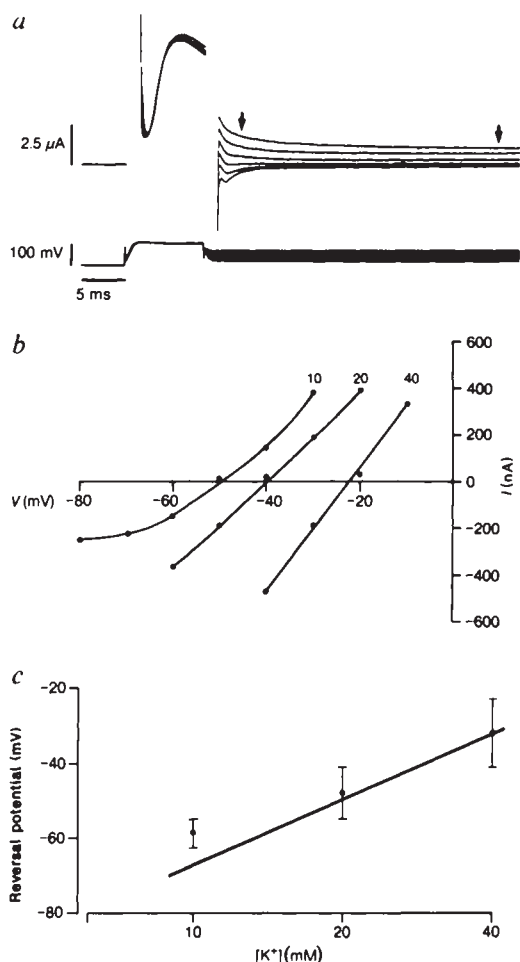


Fig. 2 The transient outward current is carried by K^+ . *a*, Examples of tail currents, and their reversal in 20 mM KCl (upper traces). The membrane was stepped from a holding potential of -80 mV to $+20$ mV, then returned to various test potentials between -60 and -10 mV (lower traces). Linear leakage current has been subtracted. *b*, Reversal of the tail current as a function of extracellular K^+ concentration. Currents were measured at about 4 and 32 ms after the second step (arrows), and the difference was plotted against the test potential. The K^+ concentrations are noted (in mM) near the current-voltage curves. Measurements were made on the oocyte of part *a*. *c*, Average reversal potentials from 4 (10 mM K^+) or 3 (20 and 40 mM K^+) oocytes $\pm$ s.e.m. are plotted against extracellular K^+ concentration. The line has the slope predicted by the Nernst equation (17 mV per twofold change in K^+ concentration at 12°C), and is drawn through the point at 40 mM K^+ . The concentration of KCl was raised at the expense of NaCl, so that the Cl^- concentration remained constant.

splicing^{6,7,16}. To test the function of *Shaker* products, we have used the *Xenopus* oocyte expression system¹⁷.

Expression in oocytes

The ShA1 and ShB1 complementary DNA clones arise from alternative splicing¹⁶. Both contain complete open-reading frames^{6,16}. The predicted proteins both include six hydrophobic sequences that may span the membrane, and an arginine-rich S4-like sequence similar to those found in the voltage-sensitive sodium channel¹⁴ and in the dihydropyridine receptor¹⁵ (Fig. 1*a*). We subcloned the ShA1 and ShB1 cDNA clones into a plasmid expression vector, transcribed *Shaker* mRNA *in vitro*, and injected each type of mRNA into a separate group of *Xenopus* oocytes. Figure 1 shows that oocytes injected with either *Shaker* mRNA displayed transient outward currents when

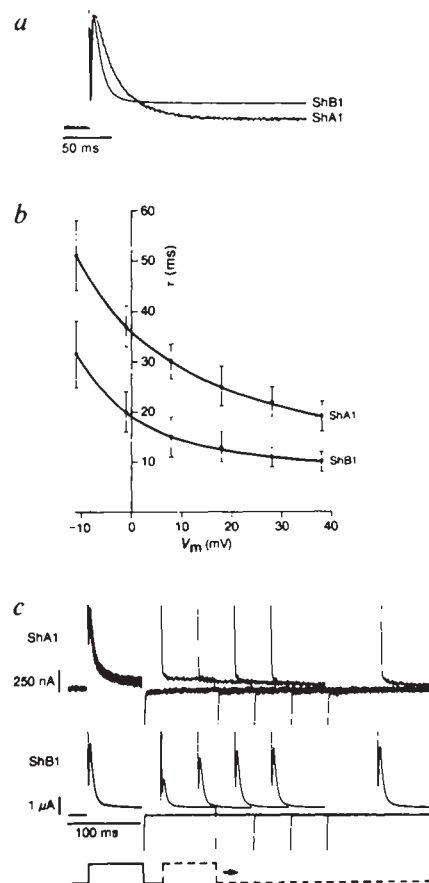


Fig. 3 A-currents induced by ShA1 and ShB1 mRNA differ in their inactivation kinetics. *a*, ShA1 and ShB1 currents evoked by steps to $+40$ mV. The currents have been scaled so that the peaks coincide. Peak amplitudes were 420 nA for ShA1 and 6.3 μ A for ShB1. The fast component of inactivation is more rapid in the ShB1 current, but inactivation is still incomplete after 225 ms. *b*, The time constants (T) of fast inactivation components $\pm$ s.e.m., plotted as a function of membrane potential. Each point is the average of measurements on 8–12 oocytes. Time constants were measured by displaying the logarithm of the current trace, identifying an early, linear portion by eye, and measuring its slope. *c*, The time course of recovery from inactivation at -80 mV, as revealed by double-pulse experiments. The membrane potential was stepped from -80 mV to $+20$ mV, returned to -80 mV, then stepped a second time to $+20$ mV. The interval between the two steps varied (indicator trace at bottom). Each current record shows five superimposed traces, in which the interval between steps varied from 25 ms to 350 ms. The A current induced by ShA1 did not recover significantly from inactivation for several minutes.

examined under voltage clamp. Of the oocytes that survived the injection and were healthy enough to study, all had transient outward currents (15 oocytes for ShA1 and 12 for ShB1). In contrast, control oocytes injected with a truncated form of ShB1 mRNA lacked the transient outward current (five of five oocytes; Fig. 1*c*). The template for the control *Shaker* mRNA was derived from ShB1 cDNA, except that the 3' end was shortened, so that it contained sequences for the first 573 amino acids of the 656 predicted from the intact ShB1 cDNA clone. This control shows that the transient outward currents did not arise from artefacts in the procedures of *in vitro* transcription and oocyte injection, but rather from the presence of intact *Shaker* mRNA.

To determine whether this transient outward current is a potassium current with the properties of the A current, and to compare it to the current in fly muscle that depends on the

Shaker product, we examined the voltage dependence of activation, rate of inactivation, ionic selectivity and sensitivity to 4-aminopyridine (4-AP) of the current in oocytes. As shown in Fig. 1, the transient outward currents first appeared as the potential of the oocyte membrane was stepped from -80 mV to between -40 mV and -30 mV. A similar threshold for A-channel activation has been found in fly larval muscle⁹, pupal muscle^{10,18} and embryonic myocyte¹¹. When the rate of channel inactivation was measured by fitting the decline of the transient current with a single exponential function, the time constants of the currents induced by ShA1 and ShB1 mRNA were about 20 ms and 10 ms, respectively, at $+40$ mV (Fig. 3b). These values are comparable to the value of 18 ms determined previously in pupal muscle¹⁰.

If the transient outward current in oocytes is carried by K^+ , the reversal potential should depend on the extracellular concentration of K^+ . In oocytes injected with ShB1 mRNA the reversal potential, determined from tail currents, was -59 mV in saline with 10 mM K^+ (Fig. 2). The reversal potential shifted up by 11 mV as the extracellular potassium concentration was raised from 10 mM to 20 mM, and by a further 16 mV as the concentration was raised from 20 mM to 40 mM. The Nernst equation predicts a 17 mV shift for a twofold increase in extracellular potassium. The difference between the observed 11 mV shift at low K^+ concentration and the predicted value may be due to experimental error: in 10 mM K^+ the tail currents are small and their kinetics are rapid, making the reversal potential difficult to measure accurately. It is also possible that the channel is slightly permeable to another ion, in addition to K^+ . In any case, the observed dependence of reversal potential on K^+ concentration shows that the channels induced by *Shaker* mRNA are permeable to K^+ .

To compare the pharmacology of the transient outward currents in oocytes to that of known A currents, we applied 4-AP, which blocks A-current selectively in molluscan neurons¹⁹. At 5 mM, 4-AP reduced the peak current amplitude in oocytes by an average of 70% (3 ShB1 oocytes; 1 ShA1 oocyte), which is comparable to the reduction of $\sim 80\%$ found in *Tritonia* neurons¹⁹. In each of these experiments the effects of 4-AP reversed partially after washing in normal saline for 15 min. Thus the channels expressed in frog oocytes injected with *Shaker* mRNA were A channels as judged by voltage sensitivity, inactivation kinetics, ionic selectivity and sensitivity to 4-AP. These results show that in *Xenopus* oocytes the *Shaker* product alone suffices to form a functional A channel.

As each *Shaker* product resembles one of four mutually similar domains in the α subunit of the sodium channel and in the dihydropyridine receptor^{14,15}, it is tempting to speculate that the *Shaker* product is a subunit of the A channel⁶. That a single *Shaker* mRNA species can generate functional A channels suggests that the channel has a homomultimeric structure in oocytes. Although the structure of the A channel in the fly is not known, the observation of allele-specific negative complementation between two *Shaker* mutations has been interpreted to mean that the A channel of pupal flight muscle contains more than one copy of the *Shaker* gene product¹⁰. It is possible that in the fly the channel contains heterologous subunits, for example some combination of the *Shaker* A, B, C or D proteins. The question of subunit structure must be addressed by biochemical

characterization of the channels purified from frog oocytes and from the fly. Nevertheless, the finding that injection of a single mRNA species produced functional A channels in oocytes will facilitate structure/function studies. A comparison of the ShA1 and ShB1 currents, for example, has begun to identify portions of the gene products that influence one kinetic process.

Inactivation kinetics

Although both ShA1 and ShB1 mRNA induced A currents in the oocytes, the kinetic properties of the two A currents differed. At every potential between -10 mV and $+40$ mV the ShA1 current inactivated more slowly than did the ShB1 current (Fig. 3a and b). Furthermore the inactivation of the ShA1 current was usually complete after several hundred milliseconds, whereas the ShB1 current did not inactivate completely (Fig. 3a). Another difference appeared when we examined the rate of recovery from inactivation at -80 mV: whereas the ShB1 current recovered almost completely within 350 msec, the ShA1 current showed no indication of recovery (Fig. 3c). Typically the ShA1 current required minutes to recover from inactivation when the membrane potential was held at -80 mV; the recovery could be hastened by repeated hyperpolarizing steps to -120 mV. Thus in *Xenopus* oocytes the two classes of *Shaker* mRNA form two types of A channels with different kinetic properties. The A current of pupal muscle resembles the ShB1 rather than the ShA1 current, with a time constant of about 650 ms for recovery from inactivation at -80 mV¹⁸. The comparison is indirect, however, as the measurements in oocytes were made at 12°C , whereas those in pupal muscle were made at 4°C .

The predicted proteins of ShA1 (616 amino acids) and ShB1 (656 amino acids) are identical from the amino terminus through most of the proposed membrane-spanning regions, including the S4-like sequence, but differ in sequence in the carboxy-terminal third¹⁶ (Fig. 1a). Even where the two cDNA clones are encoded by different exons, they continue to show considerable sequence similarity at the amino-acid level¹⁶. The different inactivation kinetics must depend on a function supplied by the non-identical residues in this region of the molecule.

Although physiological studies of the A current affected by *Shaker* mutations have been limited to muscles and myocytes thus far, molecular studies of *Shaker* have revealed potential diversity. Alternative splicing generates multiple *Shaker* products; their differential distribution in the fly suggests that they may form subtypes of A channels with different tissue distributions¹⁶. The different inactivation kinetics of the ShA1 and ShB1 currents suggest that these subtypes may have different physiological properties. It is important to determine the subunit structure of the A channel in the fly, as it is possible that those A channels contain more than one *Shaker* product. The number of possible heteromultimeric complexes made from the known *Shaker* products is quite large. These potential channels could differ in their tissue distribution, developmental expression, or in their physiological properties.

We thank Dr D. Yoshikami for helping with some preliminary experiments, Drs M. Cyert and M. Kirschner for the frogs, and Ms L. Schulte for help in preparing the manuscript. This study was supported by the NIH, a Klingenstein Fellowship and the Howard Hughes Medical Institute.

Received 29 October; accepted 18 November 1987.

- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, 1984).
- Shuster, M. J., Camardo, J. S., Siegelbaum, S. A. & Kandel, E. R. *Nature* **313**, 392–395 (1985).
- Farley, J. & Auerbach, S. *Nature* **319**, 220–223 (1986).
- Kamb, A., Iverson, L. E. & Tanouye, M. A. *Cell* **50**, 405–413 (1987).
- Papazian, D. M., Schwarz, T. L., Tempel, B. L., Jan, Y. N. & Jan, L. Y. *Science* **237**, 749–753 (1987).
- Tempel, B. L., Papazian, D. M., Schwarz, T. L., Jan, Y. N. & Jan, L. Y. *Science* **237**, 770–775 (1987).
- Baumann, A. et al. *EMBO J.* **6**, 3419–3430 (1987).
- Salkoff, L. & Wyman, R. *Nature* **293**, 228–230 (1981).
- Wu, C.-F. & Haugland, F. J. *Neurosci.* **10**, 2626–2640 (1985).

- Timpe, L. C. & Jan, L. Y. *J. Neurosci.* **7**, 1307–1317 (1987).
- Sole, C. K., Zagotta, W. N. & Aldrich, R. W. *Science* **236**, 1094–1098 (1987).
- Connor, J. A. & Stevens, C. F. *J. Physiol.* **213**, 21–30 (1971).
- Neher, E. *J. gen. Physiol.* **58**, 36–53 (1971).
- Noda, M. et al. *Nature* **312**, 121–127 (1984).
- Tanabe, T. et al. *Nature* **328**, 313–318 (1987).
- Schwarz, T. L., Tempel, B. L., Papazian, D. M., Jan, Y. N. & Jan, L. Y. *Nature* **331**, 137–142 (1988).
- Gurdon, J. B. & Wickens, M. P. *Meth. Enzym.* **101**, 370–386 (1983).
- Salkoff, L. & Wyman, R. *J. Physiol.* **337**, 687–709 (1983).
- Thompson, S. H. *J. Physiol.* **265**, 465–488 (1977).
- Melton, W. A. et al. *Nucleic Acids Res.* **12**, 7035–7057 (1984).
- Contreras, R., Cheroute, H. & Fiers, W. et al. *Analyt. Biochem.* **113**, 185–187 (1981).

Linear polarized fluctuations in the cosmic microwave background

R. B. Partridge*, Jean Nowakowski† & H. M. Martin‡

* Haverford College, Haverford, Pennsylvania 19041, USA

† Bryn Mawr College, Bryn Mawr, Pennsylvania 19010, USA

‡ Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA

It has been recognized since the early work of Silk¹ and Sachs and Wolfe² that inhomogeneities in density, velocity fields or gravitational potential can produce small amplitude fluctuations in the intensity of the cosmic microwave background on small angular scales. Searches³ for such intensity fluctuations on scales of arc seconds to degrees have now reached sensitivities of 10^{-4} in $\Delta T/T$ or better (where T is the temperature of the background). It is less often noted that density inhomogeneities can introduce small-scale fluctuations in linear polarization into the microwave background. Work by Bond and Efstathiou⁴ in the context of cold dark matter models for the origin of galaxies, for instance, suggests that fluctuations in polarization will be present on characteristically smaller angular scales than total intensity fluctuations, and will be $\sim 1/10$ their amplitude. Lubin *et al.*⁵ have shown that the large-scale linear polarization is ≤ 0.2 mK, or $\leq 7 \times 10^{-5}$ in $\Delta T/T$. We report here limits on the linear (and circular) polarization of the cosmic microwave background on small angular scales, $18'' \leq \theta \leq 160''$.

As part of a programme⁶ to set limits on fluctuations in the cosmic microwave background (CMB) on arc minute and subarc minute scales using aperture synthesis, we used the Very Large Array (VLA) of the US National Radio Astronomy Observatory to map a 43^2 arc min² portion of the sky at wavelength $\lambda = 6$ cm. The observations were made in September 1984 and are described by Martin and Partridge⁶. Because each antenna of the VLA has both a left-polarized and a right-polarized feed, it is possible to make independent maps of the sky in each of the four Stokes parameters. We have described elsewhere⁶ limits on $\Delta T/T$ resulting from the total intensity (Stokes parameter I) map. Here we describe limits on possible polarized fluctuations in the CMB which result from our maps in Stokes parameters Q and U (linear polarization) and V (circular polarization). Because the polarized emission from most radio sources is small, and because some instrumental effects are reduced in polarized maps, limits on the fractional polarization $\Delta T_p/T$ are somewhat tighter than limits on fluctuations in total intensity on the same angular scales.

We mapped a field centred at 3 h 10 min 00 s right ascension and $80^\circ 08'$ declination known to be free of radio sources brighter than ~ 3 mJy at 6 cm. Our observations lasted 25 h. Slightly less than 5% of the data, largely from noisy antennas or correlators, was dropped. We then applied a Fourier transform to the remaining visibility data using natural weighting, in which the gridded $U-V$ data are weighted according to the density of data points in the $U-V$ plane.

Because the discrete sources in the field were barely visible in the polarized maps, we cleaned⁷ the images only slightly, using 25 clean components. The clean components were not restored to the map; a separate analysis of maps with the cleaned 'sources' restored showed little difference in the resulting limits on $\Delta T/T$ (less than one standard deviation). Five arc second pixels were used for these maps; the full width of the synthesized beam of the array was $\theta = 18$ arc s. We also produced and cleaned images with $36''$ and $60''$ resolution by tapering the $U-V$ data. The pixel size in these images was 10 arc s and 16 arc s, respectively.

To distinguish between instrumental noise and real polarized fluctuations in the sky, we make use of the different angular

Table 1 Upper limits on polarized fluctuations in the CMB, at 95% confidence

Range of angular scales	Stokes parameter	Instrumental variance σ_I^2 ($\mu\text{Jy}/\text{beam}$) ²	Sky variance σ_s^2 ($\mu\text{Jy}/\text{beam}$) ²	Upper limit on $\Delta T_p/T \times 10^4$
18''–160''	Q	66.1 ± 1.2	-3.3 ± 5.7	1.4
18''–160''	U	65.1 ± 1.1	5.1 ± 4.9	2.1
18''–160''	V	66.3 ± 1.5	2.6 ± 7.3	2.2
36''–160''	Q	102.6 ± 3.8	-10.7 ± 18.8	0.66
36''–160''	U	95.8 ± 3.5	15.9 ± 15.6	0.94
36''–160''	V	107.4 ± 4.6	13.2 ± 22.5	1.04
60''–160''	Q	165.1 ± 8.9	-8.2 ± 43.9	0.42
60''–160''	U	152.6 ± 8.2	13.4 ± 36.1	0.45
60''–160''	V	183.6 ± 12.0	48.7 ± 58.8	0.63

dependences of these effects^{6,8,9}. Instrumental and atmospheric noise appears in the visibility data with random phase, so we assume it is uniformly distributed over the maps. Any polarized fluctuations of astronomical origin, on the other hand, are multiplied by the primary power pattern of the individual telescopes of the array, and hence peak at the map centre. Following ref. 6, we divide the inner 21^2 arc min² of each map into 64 square blocks and compute the variance σ^2 of the measurements in each block. The measured variance may be expressed as the sum of two components, instrumental and atmospheric variance σ_I^2 , and real sky variance σ_s^2 , which is weighted by the square of the power pattern:

$$\sigma^2 = \sigma_I^2 + \sigma_s^2 P^2(\theta)$$

where $P(\theta)$ is the primary power pattern, which we approximate as a gaussian of full width at half power 9.06 arc min. An unweighted least squares fit as a function of P^2 gives the values of σ_I^2 and σ_s^2 tabulated below. This method gives the sky variance on all angular scales between the resolution limit and the size of the block used in the analysis, which was 160 arc s. No correction was made for the presence of polarized emission from radio sources. In any case, the sources in the field⁶ were barely visible. To see what effect polarized emission from the sources had, we eliminated one or two 160 arc s square blocks containing excess flux (out of the total of 64) before making the least squares fit; the effect on the results tabulated ranged from nearly zero to about one standard deviation. In the majority of these trials, eliminating such blocks decreased σ_s^2 , but in about one third of the trials σ_s^2 increased. In Table 1 we give the results of the least squares fit with these blocks eliminated.

In most cases, the values of σ_s^2 are consistent with zero: we find no significant evidence of polarized fluctuations in the sky. Our values of σ_s^2 may be converted to upper limits on the polarized fluctuation in the CMB by

$$\Delta T_p/T = \frac{\sigma_s \lambda^2}{2kT\Omega}$$

where Ω is the synthesized beam solid angle, taken as $1.13\theta^2$, and we take $T = 2.75$ K. The resulting upper limits on polarization on the indicated scales are shown in Table 1. We present 95% confidence level upper limits, calculated by taking the given value of σ_s^2 plus 1.645 times its error. These upper limits are about three times lower than upper limits on the total intensity fluctuations set on the same angular scales⁶. Our upper limits at $\theta = 60$ –160 arc s are also comparable to the best upper limits on total intensity fluctuations set on larger angular scales using filled apertures³.

The sensitivity we achieved was limited essentially by the instrument noise in our 25-h integration. Increasing the sensitivity will therefore require either more integrating time or lower noise receivers or both. We note also that an array with smaller baselines and a larger primary beam pattern would allow measurements on angular scales 1–10 arc min, better matched to the theoretical predictions of Bond and Efstathiou⁴.

This work was supported in part by a grant from the Dudley Observatory. R.B.P. thanks George Efstathiou for piquing his interest in this problem, and Tod Ramseyer of Haverford College for help with the analysis of the data.

Received 12 October; accepted 18 November 1987.

1. Silk, J. *Astrophys. J.* **151**, 459–471 (1968).
2. Sachs, R. K. & Wolfe, A. M. *Astrophys. J.* **147**, 73–90 (1967).
3. Uson, J. & Wilkinson, D. T. *Nature* **312**, 427–429 (1984).
4. Bond, J. R. & Efstathiou, G., *Mon. Not. R. astr. Soc.* **226**, 655–687 (1987).
5. Lubin, P., Melese, P. & Smoot, G., *Astrophys. J.* **273**, L51–L54 (1983).
6. Martin, H. M. & Partridge, R. B. *Astrophys. J.* (in the press).
7. Clark, B. G. *Astr. Astrophys.* **89**, 377–399 (1980).
8. Martin, H. M., Partridge, R. B. & Rood, R. T. *Astrophys. J.* **240**, L79–L82 (1980).
9. Knoke, J. E., Partridge, R. B., Ratner, M. I. & Shapiro, I. I. *Astrophys. J.* **284**, 479–490 (1984).

A systematic asymmetry in the polarization properties of double radio sources with one jet

S. T. Garrington*, J. P. Leahy*, R. G. Conway* & R. A. Laing†

* University of Manchester, Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

† Royal Greenwich Observatory, Herstmonceux Castle, Hailsham, East Sussex BN27 1RP, UK

It has long been a puzzle why high-luminosity extragalactic radio sources, which are symmetric in many respects, have jets on one side only¹. The jets may be intrinsically asymmetric, because dissipation is stronger on one side or because they flip alternately between the two lobes they supply², or the one-sidedness may be only apparent, the result of Doppler beaming in twin relativistic jets³. So far, attempts to distinguish observationally between these two possibilities have been inconclusive⁴. Following a suggestion by Laing⁵, we report here new observations which establish that almost invariably the jet side depolarizes less rapidly with increasing wavelength than the opposite side. If the jet one-sidedness is intrinsic, this depolarization most probably occurs internally, but if the one-sidedness is due to Doppler beaming the depolarization must occur in a foreground screen. Further detailed polarization observations, which can determine whether the depolarization in such sources is internal or external, should therefore settle whether jets are intrinsically or only apparently one-sided.

Differences in the polarization characteristics of the components of double sources have been noted before⁶, but the connection with jet-sidedness only becomes apparent when high-resolution maps are compared with the polarization data⁵. To establish that there is a statistically significant effect, we have observed 55 double sources in which only one jet is seen. The selection was made without reference to polarization characteristics and was based on radio morphology, using the highest resolution maps available in the literature, supplemented in some cases by unpublished data. In most cases the jet fulfils the strict criteria of Bridle and Perley¹; in other cases the evidence for a jet is a one-sided extension from the core. Details and references are given in Table 1.

We discuss here the 25 smallest sources with angular sizes <30 arc s; analysis of the remainder will be presented elsewhere. All 25 sources are high-luminosity objects of Fanaroff–Riley⁷ Class II, with redshifts typically between 1 and 2; they include 23 quasars and 2 radio galaxies (3C41 and 3C352).

The observations were made with the National Radio Astronomy Observatory Very Large Array⁸ between May and August 1986, using the A-configuration at 1,490 MHz (wavelength λ 20 cm) and the B-configuration at 4,860 MHz (λ 6 cm) giving the same angular resolution, ~ 1.3 arc s, at both wavelengths. From the images in the Stokes parameters I , Q

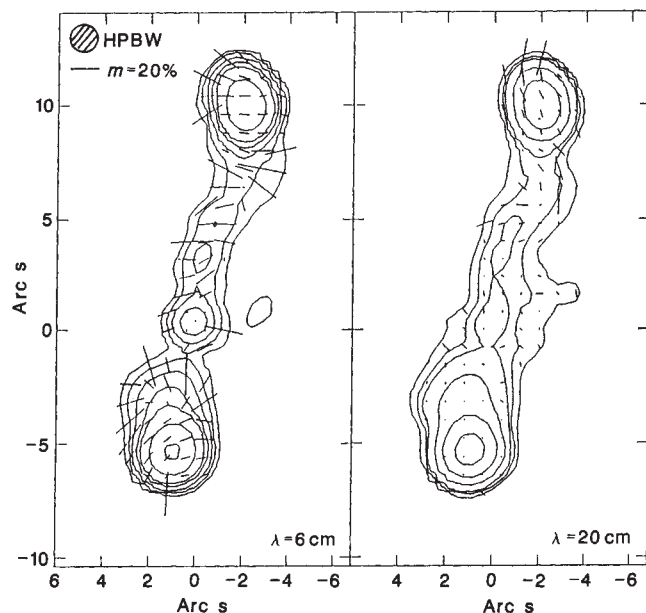


Fig. 1 Maps of 4C16.49 (1732+16) at 6 cm (left) and 20 cm (right). Contours give total intensity; the bars give the direction of the E-field of the polarized radiation, with their lengths proportional to the fractional polarization. The half power beamwidth (HPBW) is 1.3 arc s and coordinates are given relative to the position of the optical quasar. The jet points to the northern component; marked depolarization occurs on the south side, but not on the north side of the source.

and U , maps of the polarized flux density p , the E-vector position angle χ , and the fractional polarization m , were obtained. Figure 1 shows a typical example.

To quantify the depolarization effect, the maps were first masked to exclude both the flat-spectrum core and areas of low intensity ($I_{20} < 5\sigma$). Within the remaining areas the values of intensity I and polarized flux density p were summed for each side and used to calculate the quantity $m' = 100\% \times \Sigma p / \Sigma I$. This quantity, which represents the arithmetic mean percentage polarization weighted by the total intensity distribution, avoids problems arising from spurious large fractional polarizations found at the edge of a source. The values of m'_{20} , m'_6 , and of the ratio $DP = m'_{20}/m'_6$ for each side are given in Table 1.

Thermal noise, together with any residual map errors, lead to maximum errors of about $\pm 0.5\%$ in m'_{20} and m'_6 . The resulting errors in DP are usually between ± 0.05 and 0.1. Values of $m' \leq 0.5\%$ are affected by the upwards bias from noise⁹ and such values in Table 1 indicate that the component is essentially unpolarized. Apart from these cases, the noise bias is not significant and corrections for the effect have not been included.

The results listed in Table 1 and shown in Fig. 2 display a very marked asymmetry. The median DP on the jet side is 0.81, whereas the median value on the counter-jet side is 0.18. Figure 2 shows that 22 of the 25 sources in Table 1 had DP values significantly higher on the jet side. The three exceptions (3C41, 3C191 and 4C49.22) apparently depolarize at wavelengths outside our range. Longer wavelength data¹⁰ confirm that depolarization occurs in 3C41 at $\lambda > 20$ cm, and is less severe on the jet side, as in the majority. The exceptionally low degree of polarization on the counter-jet side of 3C191 suggests that this component depolarizes at $\lambda < 6$ cm. If so, this source also shows the polarization asymmetry, to an extreme degree. In summary, at least 23, and possibly 24 out of the 25 sources show less depolarization on the jet side, and there are no clear counter-examples. The probability is negligible that this asymmetry could happen by chance.

While DP gives most weight to the bright hotspots, the asymmetry is also evident in the diffuse lobe emission. DP is not correlated with the size or brightness of the components. It is known that depolarization may become stronger near the core¹⁰. The fact that in the majority (17) of our sources the peak of the emission on the jet side is further from the core than it is on the other side may therefore have influenced our result. However, this cannot be the major cause as those sources in which the component on the jet side is closer show the asymmetry just as clearly as the others. (The asymmetry in separation is not general¹¹, and is probably a selection effect: jets are hard to see if the component is close to the core). In all, the asymmetry in DP is so marked that it cannot be a secondary effect due to other differences but is genuinely connected with jet-sidedness.

Almost certainly, the depolarization we observe is due to differential Faraday rotation¹², and a greater spread in rotation is seen on the counter-jet side in some maps. We have not quoted rotation measures for the components because the values are ambiguous with only two wavelengths and also because of the low polarized signal on the counter-jet side. We can define a Faraday dispersion Δ as the standard deviation of the Faraday dispersion function¹² $F(\Phi)$, which denotes the fraction of the polarized emission coming from a Faraday depth $\Phi \propto \int n_e \mathbf{B} \cdot d\mathbf{l}$ (n_e is the thermal electron density and $\mathbf{B}$ is the magnetic field strength). Assuming a gaussian form for $F(\Phi)$ we derive from the DP ratios typical values for Δ of $60 \text{ cm}^{-3} \mu\text{G pc}$ for the jet side and $180 \text{ cm}^{-3} \mu\text{G pc}$ for the counterjet side, when corrected for the redshift of the central object.

The thermal plasma responsible for the depolarization may in principle occur either within the radio components (internal depolarization) or in an irregular foreground screen (external depolarization). The two different explanations of the jet one-sidedness make opposite predictions as to which location is the more important.

On the relativistic beaming hypothesis one expects the two lobes to be intrinsically similar. Asymmetric depolarization must then occur in a foreground screen. Since the visible jet always

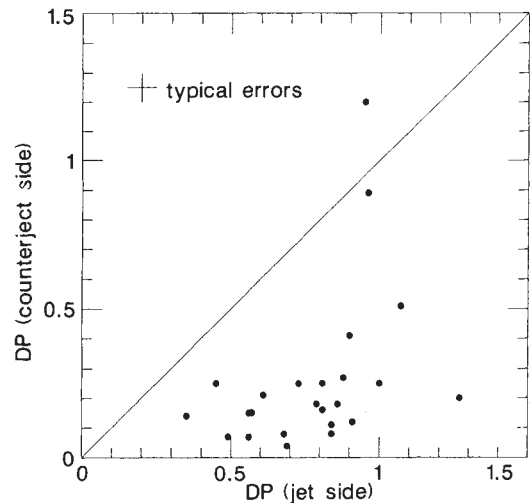


Fig. 2 Plot of depolarization ratio DP for jet side and counterjet components of 24 sources (omitting 3C191). The cross gives typical errors. For all except two sources (3C41 and 4C49.22) the values of DP are markedly higher on the jet side.

points towards us, the counterjet side is always seen through more of this screen and hence has a higher Δ . Possible configurations of the screen include a halo of X-ray-emitting gas around the central object¹³, an emission-line disk perpendicular to the radio axis¹⁴, or a cocoon of thermal plasma surrounding the emitting regions. It is implicit in this picture that the jet is directed at a small angle to the line of sight. If the depolarizing gas distribution were known, the asymmetry in DP could be used to place limits on the distribution of the angles between the axis of the radio source and the line of sight.

Table 1 Component polarizations of 25 double sources

IAU name/alias		Jet side	Jet side*			Counterjet side*			Map reference
			$m'_{20}/\%$	$m'_6/\%$	DP	$m'_{20}/\%$	$m'_6/\%$	DP	
0017+15	3C9	SE	12.8	21.0	0.61	0.9	4.5	0.20	26
0033+18	3C14	SE	6.1	4.8	1.27	2.2	11.0	0.20	23
0123+32	3C41	SE	10.9	11.4	0.96	13.2	14.9	0.89	23
0225−01	4C−01.11	NW	16.3	19.3	0.84	0.9	10.8	0.08	18
0232−04	4C−04.06	W	10.2	12.6	0.81	1.1	6.7	0.16	21
0802+10	3C191	SE	2.3	2.2	1.04	0.4	0.3	—	27
0838+13	3C207	E	6.9	9.5	0.72	3.3	13.1	0.25	22
0850+58	4C58.17	SE	14.9	26.0	0.57	2.7	18.4	0.14	19
1023+06	3C243	SE	5.9	13.2	0.45	1.3	5.1	0.25	18
1115+53		SW	11.7	13.6	0.86	1.7	9.5	0.18	25
1150+49	4C49.22	SW	12.3	12.9	0.95	17.8	14.8	1.20	24
1218+33	3C270.1	S	9.1	16.3	0.56	0.5	7.5	0.07	20
1226+10		SW	6.7	9.7	0.69	0.6	15.7	0.04	18
1241+16	3C275.1	NW	7.6	8.4	0.90	3.3	8.1	0.41	20
1258+40	3C280.1	SE	13.4	15.9	0.84	0.7	6.3	0.11	26
1308+18	4C18.36	SE	10.2	9.5	1.07	2.7	5.3	0.51	18
1318+11	4C11.45	SW	4.4	8.9	0.49	0.4	6.3	0.06	18
1323+65	4C65.15	SW	11.5	14.5	0.79	1.7	9.5	0.18	25
1634+17		W	7.5	8.5	0.88	2.3	8.5	0.27	18
1634+58	4C58.32	NW	12.7	15.6	0.81	2.9	11.5	0.25	25
1656+57	4C57.28	NE	5.2	14.9	0.35	<0.7	5.0	<0.14	25
1709+46	3C352	NW	8.0	14.2	0.56	1.4	9.1	0.15	23
1732+16	4C16.49	NW	6.7	9.9	0.68	0.7	8.5	0.08	23
1816+47	4C47.48	NW	5.8	6.4	0.90	1.9	15.3	0.12	18†
2209+15		NE	10.0	10.0	1.00	1.9	7.7	0.25	18

* Values below 0.5% are subject to noise bias, but indicate that the component is essentially unpolarized. No value of DP is given for the counterjet side of 3C191 which is unpolarized at both wavelengths.

† The map of 1816+47 shown in ref. 18 is inverted (P. D. Barthel, personal communication).

If, however, jets are intrinsically one-sided, differences in the two lobes will lead naturally to different internal Faraday depths, and thus cause asymmetric depolarization without the need to invoke any external Faraday screen. Because there is no systematic asymmetry in component width, the difference in Δ corresponds to a difference in the product $n_e B \sqrt{d}$, where d is the scale size of field reversals. Such a difference would arise if, for instance, the counterjet component were more subject to mixing with the external medium (greater n_e), or if the jet-side magnetic field were more tangled (smaller d).

To distinguish experimentally between these two depolarization mechanisms requires high-resolution polarization measurements. External depolarization, as required by relativistic beaming models, arises from variations in Faraday rotation across the line of sight on scales smaller than the beamwidth used. Consequently, depolarization should be less marked at higher resolution, and local values of DP should correlate with small-scale sideways gradients of Faraday rotation, but not in general with other features of the radio emission. Cygnus A and NGC6251 provide examples of this kind^{15,16}. On the other hand, internal depolarization (indicating an intrinsic asymmetry) should correlate directly with the local amount of Faraday rotation and should be greatest at the centres of the radio components, as is found in 3C66B and 3C130¹⁷.

In summary, it appears that high-resolution observations of Faraday depolarization may provide the best method of settling the controversial question of whether relativistic velocities persist in radio jets out to distances of many kiloparsecs.

We acknowledge the help of the staff at the VLA in obtaining

these observations. The National Radio Astronomy Observatory is operated by Associated Universities, under contract with the NSF. S.T.G. is supported by an SERC research studentship. We used the Jodrell Bank Starlink node for much of the data reduction.

Received 24 August; accepted 30 October 1987.

1. Bridle, A. H. & Perley, R. A. *Ann. Rev. Astr. Astrophys.* **22**, 319–358 (1984).
2. Rudnick, L. & Edgar, B. K. *Astrophys. J.* **279**, 74–85 (1984).
3. Orr, M. J. L. & Browne, I. W. A. *Mon. Not. R. astr. Soc.* **200**, 1067–1080 (1982).
4. Saikia, D. J. *Mon. Not. R. astr. Soc.* **208**, 231–238 (1984).
5. Laing, R. A. *Nature* **331**, 149–151 (1988).
6. Conway, R. G. & Strom, R. G. *Astr. Astrophys.* **146**, 392–394 (1985).
7. Fanaroff, B. L. & Riley, J. M. *Mon. Not. R. astr. Soc.* **167**, 31P–35P (1974).
8. Thompson, A. R., Clark, B. G., Wade, C. M. & Napier, P. J. *Astrophys. J. Suppl.* **44**, 151–167 (1980).
9. Wardle, J. F. C. & Kronberg, P. P. *Astrophys. J.* **194**, 249–255 (1974).
10. Strom, R. G. & Conway, R. G. *Astr. Astrophys. Suppl.* **61**, 547–571 (1985).
11. Saikia, D. J. *Mon. Not. R. astr. Soc.* **209**, 525–531 (1984).
12. Burn, B. J. *Mon. Not. R. astr. Soc.* **133**, 67–83 (1966).
13. Nulsen, P. E. J., Stewart, G. C. & Fabian, A. C. *Mon. Not. R. astr. Soc.* **208**, 185–195 (1984).
14. Fosbury, R. A. E. in *Structure and Evolution of Active Galactic Nuclei* (eds Giuricin, G. et al.) 297–308 (Reidel, Dordrecht, 1986).
15. Perley, R. A., Bridle, A. H. & Willis, A. G. *Astrophys. J. Suppl.* **54**, 291–334 (1984).
16. Dreher, J. W., Carilli, C. L. & Perley, R. A. *Astrophys. J.* **316**, 611–625 (1987).
17. Jägers, W. J. *Astr. Astrophys. Suppl.* **71**, 75–108 (1987).
18. Barthel, P. D. *thesis Univ. Leiden* (1984).
19. Barthel, P. D., Pearson, T. J., Readhead, A. C. S. & Canzian, B. J. *Astrophys. J.* **310**, L7–L10 (1986).
20. Stocke, J. T., Burns, J. O. & Christiansen, W. A. *Astrophys. J.* **299**, 799–813 (1985).
21. Hintzen, P., Ulvestad, J. & Owen, F. *Astr. J.* **88**, 709–758 (1983).
22. Hough, D. *thesis Caltech* (1986).
23. Laing, R. A., Riley, J. M. & Longair, M. S. *Mon. Not. R. astr. Soc.* **204**, 151–187 (1983).
24. Perley, R. A. in *Optical Jets in Galaxies* (eds Battrick, B. & Mort, J.) 77–81 (European Space Agency, Paris, 1981).
25. Shone, D. L. *thesis Univ. Manchester* (1985).
26. Swarup, G., Sinha, R. P. & Saikia, D. J. *Mon. Not. R. astr. Soc.* **201**, 393–400 (1982).
27. Cawthorne, T. V., Scheuer, P. A. G., Morison, I. & Muxlow, T. W. B. *Mon. Not. R. astr. Soc.* **219**, 883–893 (1986).

The sidedness of jets and depolarization in powerful extragalactic radio sources

R. A. Laing

Royal Greenwich Observatory, Herstmonceux Castle, Hailsham, East Sussex BN27 1RP, UK

The cause of one-sidedness in the jets of otherwise symmetrical extragalactic radio sources is a subject of some controversy¹. There is good evidence for relativistic bulk motion in the cores of these objects (from observations of apparent superluminal motion, rapid variability and the absence of inverse-Compton X-rays)². A natural hypothesis is then that jets are intrinsically two-sided and that they remain relativistic on all scales³, the jet closer to us appearing bright as a result of Doppler beaming. Alternatively, the jets may be intrinsically one-sided, in which case they need not be relativistic. During a study of the polarization properties of powerful radio sources, it became clear that in those sources with one-sided jets, depolarization with increasing wavelength is usually weaker for the lobe containing the jet. One obvious interpretation is that the depolarization is caused by differential Faraday rotation through irregularities in a magnetoionic medium surrounding the radio source. The side with the stronger jet is closer to us, is seen through a smaller amount of material and therefore shows less depolarization. A halo of hot gas around the associated galaxy or quasar is a likely candidate for the depolarizing medium.

The observations presented here refer entirely to powerful double radio sources of Fanaroff–Riley⁴ Class 2 (FR2), selected from the 3CR catalogue⁵. These have the characteristic morphology of two compact outer hot-spots equally spaced about the associated galaxy and joined by a bridge of diffuse emission. Radio jets in these sources are almost always one-sided, in the sense that the brightness ratio between jets on opposite sides of the nucleus is $>4:1$ (ref. 1). One set of depolarization measurements comes from a compilation of integrated component

Table 1 Integrated component polarizations

Source IAU	3C	Jet or* knot side	Jet $\lambda_{1/2}$ /cm	Counter-jet $\lambda_{1/2}$ /cm	Refs§
0033+183	14	Sf J	>21	10	A
0107+315	34	Sf K	40	19	A
0123+329	41	Sf K	50	30	A
0133+207	47	Sp J	17	18	B†
0824+294	200	Sf J	32	20	ref. 16
0903+169	215	Sf J	25	32	C
1241+166	275.1	Np J	>31	25	ref. 17
1618+177	334	Sf J	40	20	C
1622+238	336	Sp J	30	10	D
1626+278	341	Np J	35	>50	ref. 1‡
2120+168	432	Sf K	26	18	D
2203+292	441	Np K	17	40	A¶

* N, S, f and p denote North, South, following and preceding respectively. J indicates that the jet is continuous¹ and K that one or more knots have been detected.

† Synthesis observations (Table 2) give the opposite sidedness.

‡ 3C341 may have a two-sided jet (the brighter side has been chosen).

§ References: A, R.A.L. et al., in preparation; B, J. P. Leahy, personal communication; C, J. F. C. Wardle, personal communication; D, A. H. Bridle, personal communication.

polarizations at wavelengths between 13 and 49 cm (ref. 6). The FR2 sources from this list which are known to have jets or compact knots between their nuclei and one hot-spot are listed in Table 1. This gives the wavelengths, $\lambda_{1/2}$, at which the polarization falls to one half of its intrinsic (zero-wavelength) value for the jet and counter-jet sides of each source, integrating over all emission except for the core. Independent observations confirm the sidedness of the depolarization in four cases (3C14, 3C41, 3C275.1 and 3C441)⁷ and contradict it in one (3C47)⁸.

The second set of measurements comes from synthesis observations with the Cambridge 5-km and One-Mile Telescopes⁸ and the Very Large Array (VLA) (R.A.L. et al., in preparation). The VLA observations were made with the B-configuration⁹ at 4,885 MHz ($\lambda = 6$ cm) and the A-configuration at various frequencies between 1,413 and 1,465 MHz ($\lambda = 20$ cm), giving resolutions at full-width half-maximum (FWHM)

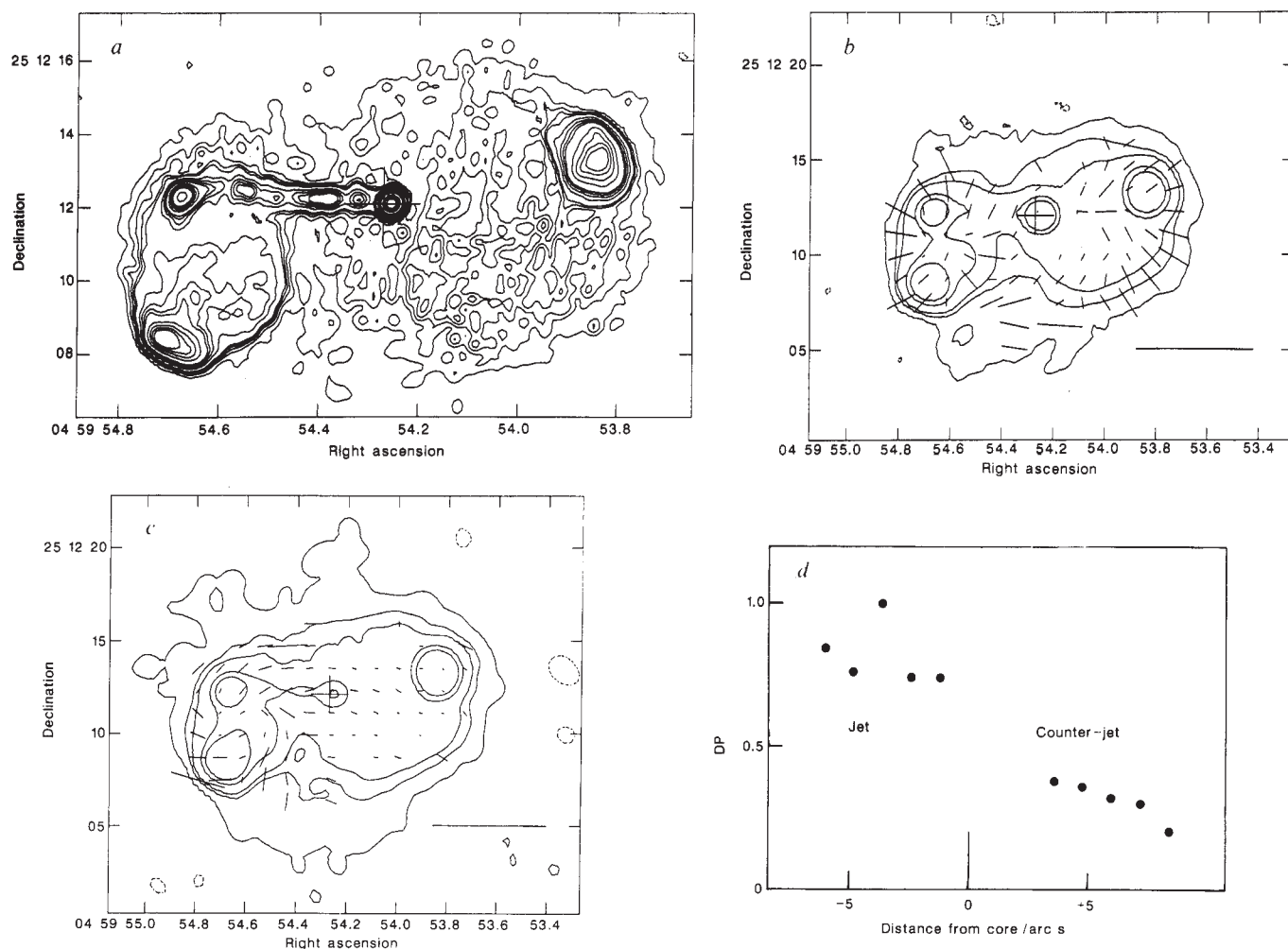


Fig. 1 Observations of the radio galaxy 3C133. On all of the maps, the cross shows the position of the associated galaxy⁵. *a*, A VLA A-configuration map at 4.9 GHz, (R. A. Laing *et al.*, in preparation) showing the jet. The beamwidth (FWHM) is 0.35 arc sec. *b* and *c*, Vectors whose lengths are proportional to the degree of polarization (measured at 4.885 GHz (*b*) and 1.465 GHz (*c*)) and whose directions are those of the *E*-vectors of linear polarization, superimposed on a few representative contours of total intensity. The horizontal line represents 100% polarization. The beamwidth is 1.15 arc s. *d*, A profile of the depolarization, DP, between 4.885 and 1.465 GHz parallel to the jet. The values of DP have been averaged in strips perpendicular to the jet.

between 1.15 and 1.3 arcs at both frequencies. The data and their reduction will be described more fully elsewhere. Maps of the degree of polarization at each frequency were generated from those of the individual Stokes' parameters, using only points at which the signal-to-noise ratio in total intensity was greater than 5, and were used to produce maps of the parameter $DP = (\text{degree of polarization at lower frequency divided by degree of polarization at higher frequency})$. Figure 1 shows an example. The mean values of DP over each lobe (excluding the region around the core) are listed for each source in Table 2.

In 9 of the 10 mapped sources, the jet side depolarizes less rapidly than the counter-jet side (Fig. 2*b*). The component polarizations give a weaker (8/12) effect in the same sense (Fig. 2*a*), but are expected to be less accurate, partly because blending with an optically thick core may affect the depolarization wavelength, but mainly because the use of a vector average of the polarized flux of a component can give misleading results, as is clearly the case for 3C47 (compare Tables 1 and 2). Mapping of an independent sample, however, confirms the result (23 out of 25 sources)⁷.

The depolarization is certainly a result of Faraday rotation in the neighbourhood of the sources. The interpretation of the result depends entirely on the location of the depolarizing material within or in front of the synchrotron-emitting plasma.

In the former case, the sidedness of the jets is probably intrinsic; otherwise, it is likely that the brighter jet is seen through a smaller amount of material and is therefore closer to us, as expected on the beaming model. The first hypothesis is explored in more detail in the companion paper⁷. Multifrequency observations at higher resolution can be used to distinguish between the possibilities^{7,10-12}.

Although no conclusive test is possible using the present data, there is one firm piece of evidence in favour of the foreground hypothesis: in the only powerful FR2 source which has been studied in sufficient detail for its depolarizing medium to be located unambiguously, Cygnus A (ref. 12), there is no doubt that the medium is in front of the radio source, rotation of the plane of polarization being $\propto \lambda^2$ over $\sim 600^\circ$, with no evidence for internal depolarization. The medium may be identified with the hot cluster gas or, less plausibly, with a denser sheath immediately around the radio lobes. The rotation measure varies from $-2,000$ to $+1,500 \text{ rad m}^{-2}$ in the western lobe (which contains the stronger jet^{13,14}) and from $< -3,000$ to $> +4,000 \text{ rad m}^{-2}$ in the eastern lobe, so the source clearly shows an asymmetry in Faraday effects in the same sense as the present samples. The X-ray observations show no gross departures from circular symmetry, so it may be assumed that the external gas density is purely a function of distance from the galaxy.

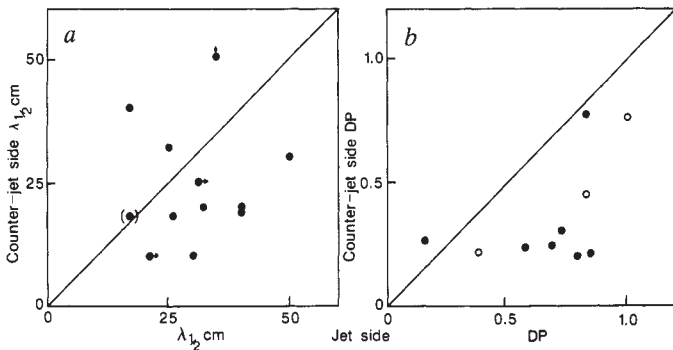


Fig. 2 Depolarization on the jet and counter-jet sides of powerful radio sources, showing *a*, integrated component polarizations from Table 1 and *b*, mean component polarizations from Table 2. Arrows denote lower limits. The bracketed point (3C47) may be incorrect (see text). In *b*, filled circles denote wavelengths of 6 and 20 cm, open circles wavelengths of 11 and 21 cm.

It is reasonable to adopt the working hypothesis that the main cause of depolarization in an FR2 source is Faraday rotation through the hot gas halo surrounding the associated galaxy, with some contribution from any denser region around the lobes and extended emission-line regions. The sources observed here must then be orientated within about 45° of the line of sight (precise constraints depend strongly on the density distribution), to generate sufficient asymmetry in path length between the two lobes, and this is consistent with the observed asymmetry in the jets. The fact that the depolarization profiles vary continuously across the source, rather than changing abruptly at the nucleus, argues for a smoothly-distributed medium rather than a thin disc. Two examples illustrate the range of physical conditions expected. In the nearby radio source M84 (ref. 15), the halo is associated with a single galaxy, whilst in Cygnus A (ref. 12) the scale is that of an entire cluster. The characteristic wavelength, $\lambda_{1/2}$, is simply related to the amplitude of rotation measure (RM) variations in the foreground screen. If the variations have a gaussian distribution with standard deviation σ , then $\lambda_{1/2} = (\ln 2/2)^{1/4} \sigma^{-1/2}$. For example, the resolved RM variations in Cygnus A and M84 have $\sigma \approx 1,000 \text{ rad m}^{-2}$ ($\lambda_{1/2} \approx 2 \text{ cm}$) and $\sigma \approx 10 \text{ rad m}^{-2}$ ($\lambda_{1/2} \approx 20 \text{ cm}$), respectively, spanning the range in Table 1.

The result arises naturally if jet sidedness is a result of Doppler beaming, and there is already some evidence in favour of the foreground hypothesis. If this can be confirmed for a few sources, then the argument that jets remain relativistic on all scales is greatly strengthened.

The National Radio Astronomy Observatory is operated by Associated Universities Inc., under contract with the National Science Foundation.

Table 2 Component polarizations from synthesis data

IAU	Source	3C	Jet or knot side	Wave-length	Mean depolarization		Refs†
					Jet	Counter-jet	
0133 + 207	47	Sp	J	11, 21	0.37	0.22	ref. 8, A
0459 + 252	133	Sf	J	6, 20	0.73	0.30	Fig. 1, B
0833 + 654	204	Np	J	6, 20	0.18	0.26	ref. 16, B
0834 + 580	205	Sp	K	6, 20	0.69	0.24	B
0850 + 140	208	Sp	J	6, 20	0.80	0.20	ref. 16, B
0855 + 143	212	Np	J	6, 20	0.58	0.24	B
0917 + 458	219	Sp	J	11, 21	0.83	0.45	ref. 8
1100 + 772	249.1*	Sf	J	6, 20	0.83	0.77	ref. 16, B
1206 + 439	268.4	Sp	K	6, 20	0.85	0.21	ref. 16, B
1833 + 326	382	Nf	J	11, 21	1.00	0.77	ref. 8, C

* The base of the jet in 3C 249.1 depolarizes rapidly, probably as a result of associated emission-line gas; the overall depolarization values are dominated by the lobes and do not reflect this.

† References: A, personal communication, J. P. Leahy; B, R.A.L. *et al.*, in preparation; C, personal communication, J. W. Dreher.

Received 25 August; accepted 30 October 1987.

1. Bridle, A. H. & Perley, R. A. *A. Rev. Astr. Astrophys.* **22**, 319–358 (1984).
2. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Rev. mod. Phys.* **56**, 255–351 (1984).
3. Orr, M. J. L. & Browne, I. W. A. *Mon. Not. R. astr. Soc.* **200**, 1067–1080 (1982).
4. Fanaroff, B. L. & Riley, J. M. *Mon. Not. R. astr. Soc.* **167**, 31P–35P (1974).
5. Laing, R. A., Riley, J. M. & Longair, M. S. *Mon. Not. R. astr. Soc.* **204**, 151–187 (1983).
6. Strom, R. G. & Conway, R. G. *Astr. Astrophys. Suppl.* **61**, 547–571 (1985).
7. Garrington, S. T., Leahy, J. P., Conway, R. G. & Laing, R. A. *Nature* **331**, 147–149 (1988).
8. Burch, S. F. *Mon. Not. R. astr. Soc.* **186**, 519–553 (1979).
9. Thompson, A. R., Clark, B. G., Wade, C. M. & Napier, P. J. *Astr. Astrophys. J. Suppl.* **44**, 151–167 (1980).
10. Burn, B. J. *Mon. Not. R. astr. Soc.* **133**, 67–83 (1966).
11. Laing, R. A. in *Physics of Energy Transport in Extragalactic Radio Sources: NRAO Workshop No. 9* (eds Bridle, A. H. & Eilek, J. A.) 90–96 (NRAO, Green Bank, 1984).
12. Dreher, J. W., Carilli, C. L. & Perley, R. A. *Astr. Astrophys. J.* **316**, 611–625 (1987).
13. Perley, R. A., Dreher, J. W. & Cowan, J. J. *Astr. Astrophys. J.* **285**, L35–L38 (1984).
14. Dreher, J. W., Carilli, C. L. & Perley, R. A. *Bull. Amer. astr. Soc.* **19**, 731 (1987).
15. Laing, R. A. & Bridle, A. H. *Mon. Not. R. astr. Soc.* **228**, 557–571 (1987).
16. Burns, J. O., Basart, J. P., De Young, D. S. & Ghiglia, D. C. *Astr. Astrophys. J.* **283**, 515–528 (1984).
17. Stocke, J. T., Burns, J. O. & Christiansen, W. A. *Astr. Astrophys. J.* **299**, 799–813 (1985).

The a.c. Josephson effect in constrictions engraved in bulk $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ and d.c. SQUID operation at 77 K

D. Robbes*, Y. Monfort*, M. Lam Chok Sing*, D. Bloyet*, J. Provost*, B. Raveau*, M. Doisy† & R. Stephan†

* ISMRa, Boulevard du Maréchal Juin, 14032 Caen Cedex, France
† Thomson Sintra ASM, Route de Ste Anne du Porzic, 29001 Brest, France

One of the most important potential applications of the new high-temperature oxide superconductors is the superconducting quantum interference device (SQUID). We have made a d.c. SQUID using Josephson junctions made by engraving small constrictions ($0.5 \times 0.25 \text{ mm}$) in bulk $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ ceramics; we report here some of the magnetic properties observed. Shapiro steps and well defined quantum interferences between intrinsic junctions have been observed at 77 K. The critical current I_c is clearly modulated by the applied magnetic field. The measured field sensitivity is $\sim 1.2 \text{ pT Hz}^{-1/2}$ and is limited by the noise of the detection system; this value is certainly an underestimate of the true sensitivity of the device. The corresponding energy resolution lies in the range 10^{-29} – $10^{-30} \text{ J Hz}^{-1}$.

We have studied the current-voltage (I - V) characteristic of the device shown in the Fig. 1 inset. The compound was routinely fabricated following the recipe in ref. 1. The sample was bonded to an ordinary glass substrate and then reduced to a thickness of 0.25 mm. Finally, a constriction was engraved. To prevent any degradation of the superconducting phase by exposure to water, as reported in ref. 2, a coating of light acrylic varnish was applied. The size of the junction leads to a critical current I_c of $\sim 10 \text{ mA}$; Higashino *et al.*³ have shown that this value is adequate for the observation of the a.c. Josephson effect. Moreover, multiple superconducting loops containing weak links are intrinsic to the ceramic material⁴. This led us to think that intrinsic loops with two or more weak links might exist, so that the d.c. Josephson effect and thus quantum interferences (d.c. SQUID behaviour) would be observable in the sample at 77 K. All of our measurements were carried out at liquid-nitrogen temperature and in a poorly magnetically shielded environment.

Figure 1a shows the I - V characteristic of a sample with $I_c = 6.7 \text{ mA}$. When microwave radiation was applied to the sample, Shapiro steps were obtained in the I - V characteristic (Fig. 1b–f). The voltage difference between the steps is $17.6 \pm 0.2 \mu\text{V}$ for 8.49-GHz microwave radiation, giving a ratio of $2.07 \pm 0.02 \mu\text{V GHz}^{-1}$ which agrees very well with the theoretical value of $h/2e = 2.068 \mu\text{V GHz}^{-1}$. These results are also in agreement with those reported in refs 3 and 5 for other types of junction.

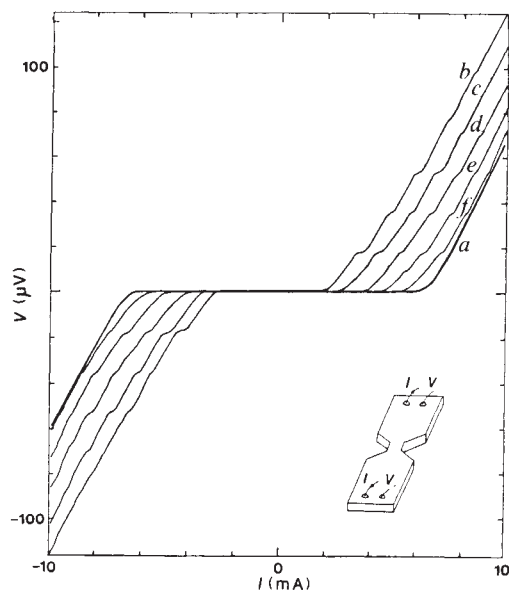


Fig. 1 Current-voltage (I - V) characteristic of a constriction ($I_c = 6.7$ mA) at 77 K for various relative levels of applied microwave power (8.49 GHz): b , 0 dB; c , -4 dB; d , -7 dB; e , -10 dB; f , -25 dB. Shapiro steps are clearly visible. Curve a is obtained when the microwave irradiation is turned off. The inset shows a schematic drawing of the constriction (0.5 mm wide, 0.25 mm thick).

I_c was then reduced both *in situ* with a current pulse technique^{6,7}, and by mechanical abrasion, and varied between 300 and 500 μ A at 77 K in subsequent coolings. These variations probably depend on intermediate thermal cycling conditions between 77 and 239 K^{8,9}. The I - V characteristic obtained for this sample (Fig. 2) exhibits hysteretic behaviour at ~ 340 μ A.

To measure precisely the magnetic field dependence of our constriction and examine the possible existence of intrinsic d.c. SQUIDs, we biased the junction with a steady current I slightly greater than I_c . We then applied a magnetic field by means of a square loop (8 $\times$ 8 mm) centred on the constriction. The voltage V across the device was measured with a differential pre-amplifier. Figure 3 shows the voltage V plotted against the applied magnetic flux ϕ at 77 K. The voltage across the device is a quasi-periodic function of the applied flux, with period ϕ_0 . We observe several tens of periods separated by unstable and/or hysteretic regions. The plot obtained is quite different from the periodic response of the d.c. SQUID as described by Clarke¹⁰, as the amplitude is somewhat field-dependent and unstable regions are observed. Nevertheless, an appropriate flux biasing renders the junction usable in the stable regions (Fig. 4).

In these conditions, we measured the voltage-to-flux transfer function $dV/d\phi$ up to 25 μ V ϕ_0^{-1} at 77 K, using a small-signal

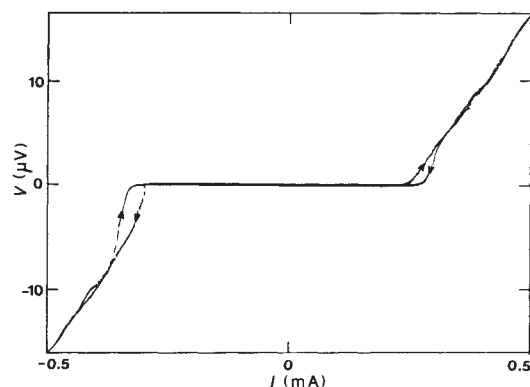


Fig. 2 I - V characteristic of a sample exhibiting d.c. SQUID behaviour. It is hysteretic at ~ 340 μ A.

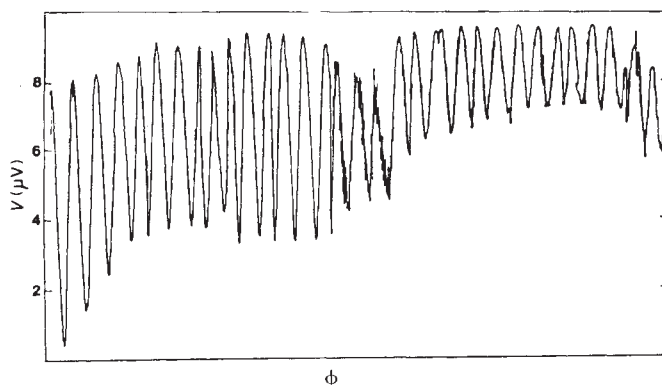


Fig. 3 Plot of the voltage V across the SQUID versus the applied flux ϕ at 77 K. The voltage reference is arbitrary. Bias current $I = 320$ μ A; bandwidth, 5 Hz.

measurement method. The noise measurements were carried out in two configurations. We first connected the device directly to the preamplifier input. The observed flux sensitivity was found to be limited by the amplifier noise (4.4 nV $\text{Hz}^{-1/2}$) to $\sim 1.8 \times 10^{-4} \phi_0 \text{ Hz}^{-1/2}$. In order to overcome this drawback, a cooled LC tank circuit (resonant frequency 27.1 kHz, Q-factor 36, resistive losses of 14Ω in the coil) was used to improve the noise matching of the SQUID impedance to the input of our pre-amplifier¹⁰. Under these conditions, we measured a flux sensitivity of $9 \times 10^{-6} \phi_0 \text{ Hz}^{-1/2}$, equivalent to a device noise voltage of 0.23 nV $\text{Hz}^{-1/2}$. We further showed experimentally that this noise was still due mostly to the resistive losses of the tank circuit and preamplifier noise. We therefore conclude that the true flux sensitivity of the device is lower than $9 \times 10^{-6} \phi_0 \text{ Hz}^{-1/2}$ at 27.1 kHz. Moreover, a current of 0.94 mA in the loop produces one period in the $V(\phi)$ function, yielding a field sensitivity of 1.2 pT $\text{Hz}^{-1/2}$.

We may now estimate the energy resolution of the device. The inductance L of the SQUID can be roughly derived in two ways. Using first the $V(\phi)$ characteristic obtained by the application of a known magnetic field, we deduce an effective loop area of $1.4 \times 10^{-8} \text{ m}^2$, which corresponds, together with the 0.25 mm thickness of the constriction, to an inductance of ~ 70 pH. Alternatively, an analysis of the I - V and $V(\phi)$ characteristics (Figs 2 and 4) leads to an estimate of the current modulation depth $\Delta I_c/I_c$ of the SQUID of ~ 10 -30%. The inductance can be estimated from the expression $L \approx \phi_0/(2\Delta I_c)$, given in ref. 10 as 10-50 pH. Assuming a value of 50 pH and a flux sensitivity of $9 \times 10^{-6} \phi_0 \text{ Hz}^{-1/2}$, the corresponding energy resolution is found to be $3.5 \times 10^{-30} \text{ J Hz}^{-1}$, or 5,200 h , where h is Planck's constant.

The presence of intrinsic weak links in bulk $\text{YBa}_2\text{Cu}_3\text{O}_{7.8}$ (refs 3, 11, 12) is supported by our observation of quantum interferences in the d.c. SQUID. This implies the coexistence of a superconducting phase and a normal one, probably semi-

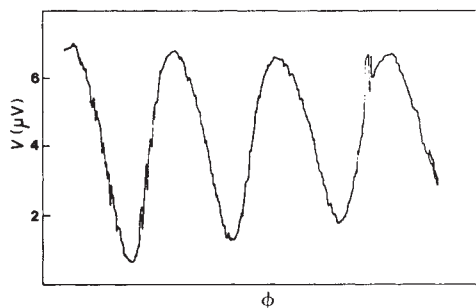


Fig. 4 V versus ϕ in the stable region where the sensitivity measurements were carried out. Temperature, bias current and bandwidth as in Fig. 3.

conducting, at 77 K. In fact, a superconducting loop interrupted by two Josephson junctions must exist for the d.c. SQUID effect to be observed. This would certainly not be the case in a perfectly homogeneous superconducting material. Moreover, we have observed a semiconductor-like temperature dependence of the resistance (for $T > 77$ K) of constrictions smaller than 0.25×0.25 mm in the same batch. At present, the complete noise spectrum of the SQUID is not available, but these preliminary results are very promising.

We thank C. Michel and F. Deslandes for their preparation of the $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ ceramic samples.

Received 3 November; accepted 18 December 1987.

1. Michel, C. *et al. C.r. hebd. Séanc. Acad. Sci., Paris* **304**, II, 1059-1061 (1987).
2. Yan, M. F. *et al. Appl. Phys. Lett.* **51**, 532-534 (1987).
3. Higashino, Y., Takahashi, T., Kawai, T. & Naito, S. *J. appl. Phys.* **7**, L1211-L1213 (1987).
4. Colclough, M. S. *et al. Nature* **328**, 47-48 (1987).
5. Moreland, J. *et al. Appl. Phys. Lett.* **51**, 540-541 (1987).
6. Duret, D., Bernhard, P. & Zenatti, D. *Rev. scient. Instrum.* **46**, 475-477 (1975).
7. Monfort, Y., Bloyet, D., Villegier, J. C. & Pascal, D. *Revue Phys. appl.* **19**, 449-454 (1984).
8. Matthews, D. N., Bailey, A., Vaile, R. A., Russell, G. J. & Taylor, K. N. R. *Nature* **328**, 786-787 (1987).
9. Bhargava, R. N., Herko, S. P. & Osborne, W. N. *Phys. Rev. Lett.* **59**, 1468-1470 (1987).
10. Clarke, J. in *Superconductor Applications: SQUIDS and Machines* (eds Schwartz, B. B. & Foner, S. 67-124 (Plenum, New York, 1977)).
11. Deutscher, G. & Muller, K. A. *Phys. Rev. Lett.* **59**, 1745-1747 (1987).
12. Esteve, D. *et al. Europhys. Lett.* **3**, 1237-1242 (1987).

A novel deposition concept for amorphous superlattices

M. Kawasaki*, Y. Matsuzaki*, K. Fueki*, K. Nakajima†, Y. Yoshida† & H. Koinuma‡§

* Department of Industrial Chemistry, Faculty of Engineering, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 112, Japan

† Department of Applied Chemistry, Faculty of Engineering, Tokyo University, Kujirai, Kawagoe, Saitama 350, Japan

‡ The Research Laboratory of Engineering Materials, Tokyo Institute of Technology, Nagatsuda, Midori-ku, Yokohama 227, Japan

Cumulative layered structures of amorphous semiconductors have attracted much interest in recent years, as they make it possible to use quantum effects to improve semiconductor properties¹. There is, however, a serious disadvantage in using such superstructures in practical devices, because the conventional process requires mechanical alternation of source gases and evacuation of the system for each layer deposition. As an alternative to this inefficient physically controlled method, we have developed an efficient photochemically controlled method which facilitates the continuous deposition of various superlattices. In combination with two modes of excitation, at least one of which is operated in a pulsed mode, an appropriately selected pair of source gases accumulates continuously binary layers of amorphous semiconductors. As an example of verification of this novel process, an amorphous superlattice composed of an amorphous silicon carbide barrier layer and an amorphous silicon well layer was prepared from a mixture of disilane and carbon tetrafluoride.

Figure 1 shows the basic principle of our method devised for the continuous preparation of layered structures of two different amorphous films. Electrons in a plasma generally have such a broad energy distribution that they can decompose a wide variety of source gases into fragments which can in turn become precursors to amorphous film deposition. On the other hand, ultraviolet light radiated from a low-pressure mercury lamp or an excimer laser can selectively decompose those gases that have sizeable cross-sections either for the light absorption or for the energy

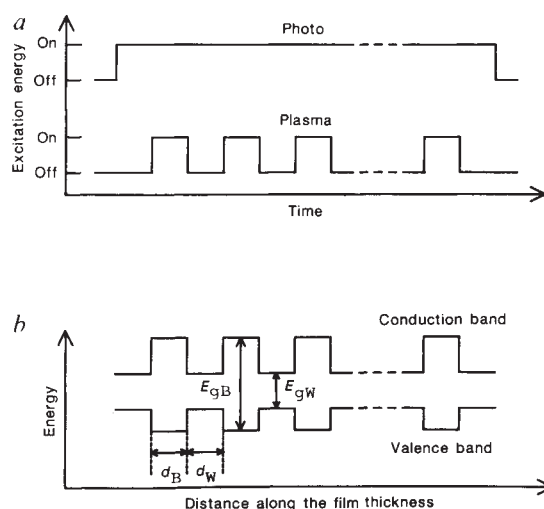


Fig. 1 Basic concept of the continuous deposition of amorphous superlattices by using differential photochemical reactivities of component source gases. *a*, The sequence of each excitation source. *b*, Band diagram of the superlattice prepared by PPP CVD in the excitation sequence mode shown in *a*. The optical band gaps of barrier layer and well layer are represented as E_{GB} and E_{GW} , respectively. The film thicknesses of barrier layer and well layer are represented as d_B and d_W , respectively.

transfer from a sensitizer, which is mixed in the reaction system and excited by the incident light. Thus, under constant irradiation from a low-pressure mercury lamp, a pulsed plasma is applied to a mixture of two types of source gases (Fig. 1a). One gas (type A) is chosen for continuous decomposition throughout the reaction, whereas the second gas (type B) should be inert for the photochemical excitation reaction during the period for which the plasma is turned off. If the decomposition of a type B gas by secondary reaction with a fragment decomposed from the type A gas is negligible, a layered structure composed of alternate A-B alloy and A films will be formed continuously. Figure 1b shows the band diagram of the resultant layered film assumed for the case when the band gap of the A-B alloy semiconductor is wider than that of pure A. The width of each layer can be freely regulated by the pulse sequence. The interface is expected to be sharp because the lifetime of active species generated in the plasma and/or photochemical excitations are estimated, from experiments on pulsed^{2,3} and triode plasmas⁴, to be far shorter (<1 ms) than the times for 1 Å growth of photo-plasma and photo chemical vapour deposition (CVD) films, 0.7 and 1.2 s, respectively, as are indicated in Table 1.

To distinguish the chemical reactivity of the source gases to plasma excitation from that for photochemical excitation, as well as to examine the feasibility of the scheme represented in Fig. 1, an apparatus for pulsed plasma and photo CVD (abbreviated as PPP CVD here) was designed and constructed as depicted in Fig. 2. A pyrex glass cylindrical chamber of 150-mm diameter was surrounded by a three-turn copper coil which was used to generate an inductively coupled plasma by applying 13.56-MHz r.f. current. The power supply to sustain the plasma can be modulated by a pulse programmer with a rise and fall time <1 μs. The pulse width (t_{on}) was fixed to be 33.3 s and the pulse interval ($t_{on} + t_{off}$) was changed in the ranges of 44.9–323.3 s. A 110-V low-pressure mercury lamp (SEN Tokushu Kogen SUV-110V) was placed above the chamber to irradiate the reaction system through the quartz window on top of the chamber. To avoid film deposition on the window surface, the window was coated with a fluorocarbon oil and exposed to a gentle flow of a type B gas. As a sensitizer, a very small amount of mercury vapour was induced into the chamber by passing a part of the source gas mixture over a mercury reservoir immersed

§ To whom correspondence should be addressed.

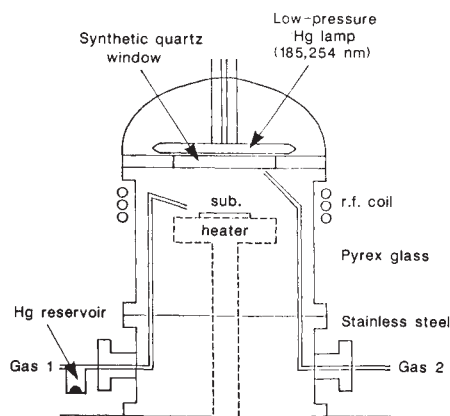


Fig. 2 Schematic representation of the PPP CVD apparatus. Gas 1: Si_2H_6 , 3.0 scm^3 ; CF_4 , 15.0 scm^3 . Gas 2: CF_4 , 45.0 scm^3 . The total flow rate of CF_4 is 60 scm^3 .

in a water bath at 60°C . Among many possible combinations of source gases, disilane was used as a type A gas and carbon tetrafluoride as a type B gas because the former is reported to have quenching cross-sections for triplet-excited mercury that are far greater than the latter^{5,6}. Typically, reactions were carried out under the following conditions: substrate temperature = 300°C , flow rates of $\text{Si}_2\text{H}_6 = 3.0 \text{ scm}^3$ (standard cubic centimetre per minute) and of $\text{CF}_4 = 60.0 \text{ scm}^3$, respectively, total pressure = 700 mtorr, and r.f. power = 30 W. The $\text{CF}_4/\text{Si}_2\text{H}_6$ ratio of 60/3 was used to make the optical band gap of the a-SiC:H:F layer wide enough compared to that of a-Si:H layer. The other conditions are as for conventional techniques.

Table 1 summarizes the conditions and results of PPP CVD of multilayers using mixtures of disilane and carbon tetrafluoride as source gases, together with those of comparative runs producing single layers. The optical band gaps and infrared spectra of films deposited from the mixed gases in the mercury-sensitized photo CVD were substantially the same as those deposited from pure disilane. This result clearly indicates that carbon tetrafluoride was hardly decomposed by collisions not only with triplet-excited mercury atoms but also with radical species generated from the photochemical decomposition of disilane. The films deposited with continuous plasma and photo-hybrid excitation, on the other hand, had optical band gaps and infrared spectra markedly different from those of the photo-CVD films. Only in the former films were observed infrared absorption

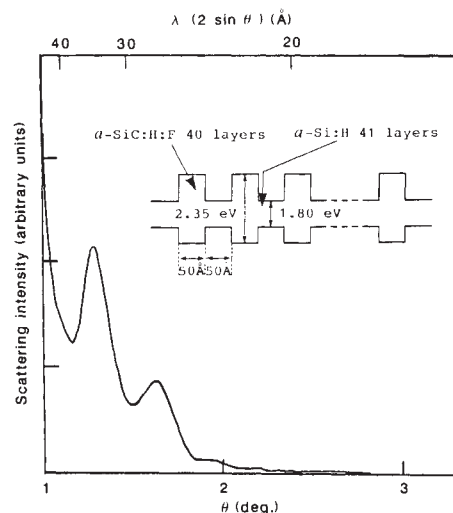


Fig. 3 X-ray diffraction pattern of an a-Si:H/a-SiC:H:F (50 Å – 50 Å , 41–40 layers) superlattice prepared in a continuous deposition by PPP CVD using disilane and carbon tetrafluoride as source gases.

bands at 900 and 770 cm^{-1} assignable to Si–F and at $1,000$ – $1,100 \text{ cm}^{-1}$ assignable to C–F stretching. Thus, our postulate that plasma-photo-hybrid CVD should give hydrogenated and fluorinated amorphous silicon carbide (a-SiC:H:F) films from mixtures of disilane and carbon tetrafluoride was confirmed. It should also be noted that carbon tetrafluoride, usually regarded as an etching gas, was used as the carbon source for the preparation of a-SiC alloys in combination with disilane, a highly reactive deposition gas.

Assuming that the thickness of a-SiC:H:F and hydrogenated amorphous silicon (a-Si:H) films prepared in this PPP CVD were proportional to the durations of plasma-photo-hybrid CVD and of photo CVD, respectively, the pulse sequence was programmed to fix the period of r.f.-on for 33.3 s and to vary the period of r.f.-off in the range between 11.6 and 291.0 s so that the thickness of a-SiC:H:F and a-Si:H layers could be regulated to be 50 Å and 10 – 250 Å , respectively. The thicknesses of the films prepared by PPP CVD were consistent with the total thicknesses designed from the deposition rate of each layer given in Table 1. Figure 3 shows the small angle X-ray diffraction pattern of a film prepared by the present PPP CVD method repeating a pulse sequence of 33.3 s on and 58.1 s off forty times. The second up to the fourth diffraction patterns were observed at angles consistent with the formation of the designed ($50 + 50 \text{ Å}$) layer structure. Depth profiles of Si, C and F atoms measured by X-ray photoelectron spectroscopy also confirmed the layered structure in the film as Si concentration showed a periodical change with reversed intensity from those of C and F concentrations. As has been already reported for the a-SiC:H/a-Si:H superlattices prepared by batch processes⁷, the films prepared by PPP CVD showed increases in their optical gaps from 1.88 to 2.15 eV as the a-Si:H layer thickness was reduced from 250 to 10 Å . Studies on further details of the semiconductor properties of these films, as well as on the application of this method to the synthesis of other amorphous superlattices, are in progress.

This study was partly supported by the Nissan Science Foundation. Disilane was supplied by Mitsui Toatsu Chemicals Inc.

Received 4 August; accepted 30 October 1987.

Table 1 Deposition conditions and optical bandgaps and deposition rates of amorphous superlattices prepared by photo-CVD, photo-plasma-CVD and PPP CVD method

Continuous CVD							
		Reaction time (min)	Thickness (Å)	Deposition rate (Å s ⁻¹)	E_g (eV)		
Photo CVD		60	3,100	0.86	1.80		
Photo-plasma CVD		30	2,700	1.50	2.35		
PPP CVD							
d_w (Å)	d_b (Å)	t_{off} (s)	t_{on} (s)	Repeat times (off-on)	Total thickness (Å)	E_g (eV)	
10	50	11.6	33.3	21–20	1250	2.15	
25	50	29.0	33.3	21–20	1550	1.97	
50	50	58.0	33.3	11–10	1050	1.96	
100	50	116.0	33.3	11–10	1550	1.96	
250	50	290.0	33.3	6–5	1550	1.88	

E_g , The optical bandgap of the film determined from the $\sqrt{ah\nu} - h\nu$ plot using the ultraviolet-visible absorption data. d_w , The width of a-Si:H well layer; d_b , the width of a-SiC:H:F barrier layer; t_{off} , the period of plasma-off; t_{on} , the period of plasma-on.

1. Abeles, B. & Tiedje, T. in *Semiconductors and Semimetals* Vol. 21, pt C, (ed. Pankove, J. I.) 407–425 (Academic, Orlando, 1984).
2. Vinzant, J. W., Shen, M. & Bell, A. T. in *Plasma Polymerization* (eds Shen, M. & Bell, A. T.) 79–85 (American Chemical Society, Washington DC, 1979).
3. Hamasaki, T., Ueda, M., Hirose, M. & Osaka, Y. *J. non-cryst. Solids* **59/60**, 679–682 (1983).

4. Matsuda, A., Kaga, T., Tanaka, H. & Tanaka, K. *J. non-cryst. Solids* **59/60**, 687-690 (1983).
 5. Pollock, T. L., Sandhu, H. S., Jodhan, A. & Strausz, O. P. *J. Am. chem. Soc.* **94**, 1017-1024 (1973).
 6. Cvetanovic, R. J. in *Progress in Reaction Kinetics* Vol. 2 (ed. Porter, G.) 39-130 (Pergamon, Oxford, 1964).
 7. Tiedje, T., Abeles, B., Persans, P. D., Brooks, B. G. & Cody, G. D. *J. non-cryst Solids* **66**, 345-350 (1984).

Uranium in pore waters from North Atlantic (GME and Southern Nares Abyssal Plain) sediments

P. H. Santschi*, C. Bajo†, M. Mantovani‡, D. Orciuolo‡, R. E. Cranston§ & J. Bruno||

* EAWAG, Swiss Federal Institute of Water Resources and Water Pollution Control, CH-8600 Dübendorf, Switzerland

† National Cooperative for the Storage of Radioactive Waste (NAGRA), CH-5401 Baden, Switzerland

‡ EIR, Swiss Federal Institute of Reactor Research, CH-5303 Wuerenlingen, Switzerland

§ Atlantic Geoscience Centre, Bedford Institute of Oceanography, PO Box 1006, Dartmouth, Nova Scotia B2Y 4A2, Canada

|| Department of Inorganic Chemistry, The Royal Institute of Technology, S-10044 Stockholm, Sweden

A central question to be addressed by the international sub-seabed radioactive waste disposal program is the mobility of radionuclides through sediments, which are the most important barrier to the transfer to the ocean and eventually to man. The major component of high-level radioactive waste is uranium. Here we report the measurement of low uranium concentrations in composite pore-water samples from the uppermost 20–30 m of deep-sea abyssal plain sediments from the Great Meteor East and Southern Nares Abyssal Plains Area. Many values are between 0.1 and 0.5 p.p.b. (parts per 10⁹) which are the lowest uranium concentrations ever measured in the pore waters of deep-sea sediments. Our lowest value, 0.05 ± 0.01 p.p.b., is orders of magnitude lower than the predicted solubility of UO_2 or U_4O_9 , probably indicating solubility control of U(IV) by an adsorbed phase. The uranium concentrations obtained from both sites correlate closely with measured redox potentials (E_h) in the sediments. The low mobility of uranium in pore waters from turbiditic deep-sea abyssal plain sediments, which can be deduced from these measurements, has important implications for the sub-seabed disposal of high-level radioactive waste, and for marine geochemistry of uranium.

The recent literature on the behaviour of uranium in natural waters has been reviewed^{1,2}. Uranium concentrations in sub-oxic waters above $E_h = -200$ mV, have been found to be a function of carbonate, and above $E_h = -100$ mV, of the fluoride and phosphate content¹. The expected concentration limits for dissolved uranium $3.24 \mu\text{g kg}^{-1}$ for seawater at 35‰ (ref. 3) and approximately 0.04 p.p.b. in pure water at equilibrium with UO_2 (uraninite)¹. Concentrations up to 16 p.p.b. in suboxic pore waters have recently been reported⁴ suggesting remobilization of uranium.

However, recent work^{4–6} suggests that in marine surface sediments, uranium is mobile only over short distances as it is released by the oxidation of organic matter by molecular oxygen that penetrates through the sediment surface from seawater. Subsequently, uranium is immobilized by reduction to U(IV), leading to substantial uranium enrichments in reducing zones.

To further substantiate the thermodynamic constraints on the mobility of uranium in suboxic pore waters of the upper 20–30 m of turbiditic sediments of the Great Meteor East (GME, Madeira Abyssal Plain) and Southern Nares Abyssal Plain, uranium concentrations were measured by mass-spectrometric isotope-dilution methods from 10 ml composite pore water samples after extraction of U by TAO (modified from ref. 10).

Pore water was extracted from ESOPE sediment cores 24

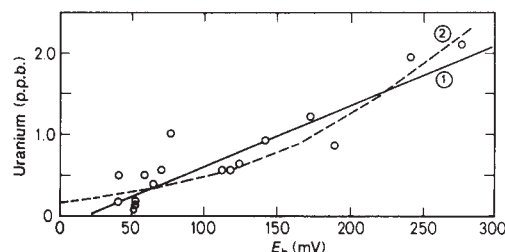


Fig. 1 Correlation of uranium concentrations in sediment pore waters with E_h (sediment). Least square fits to the data points, excluding the two extreme values (see text), give (1) $[\text{U}]$ (p.p.b.) = $0.0073 (\pm 0.0009) E_h (\text{mV}) - 0.09 (\pm 0.14)$, with $n = 17$, $r = 0.91$, $P \leq 0.0001$ or (2) $\log U$ (p.p.b.) = $0.0042 E_h (\text{mV}) - 0.75$, with $n = 17$, $r = 0.80$, $P \leq 0.001$.

Methods. Before analysis, the pore water sample was spiked with $47.4 \text{ ng } ^{233}\text{U}$ (containing 0.07% ^{238}U) and separated by liquid-liquid solvent extraction by triphenyl-arsin oxide (TAO). This separation method was developed in ref. 10 for spectrophotometric uranium determination and further adapted to mass spectrometric requirements at EIR. ^{238}U concentrations were calculated after appropriate corrections were applied for the isotopic abundance and blanks in tracer and sample solutions. Further details on uranium analysis are given in ref. 22. Because the expected concentrations are rather low, contamination is a serious problem. Blanks (tridistilled H_2O and 0.7 M NaCl solutions submitted to the chemical separation procedure) were first measured to estimate the contamination level. To check any bias, standard addition was performed on 0.7 M NaCl solutions, and seawater was analysed. The average distilled water blank (10 ml) from the determinations was $0.42 \pm 0.13 \text{ ng}$ ($n = 3$), whereas the uranium blank of 10 ml of a 0.7 M NaCl solution was 1.57 ng . A recovery test in a 10 ml 0.7 M NaCl solution, where $19.23 \text{ ng } ^{238}\text{U}$ had been added, gave $19.66 \pm 0.15 \text{ ng U}$ ($n = 3$), which is a 102.2% recovery. A seawater sample at 35‰ salinity gave a blank-corrected value of $3.39 \mu\text{g/l} \approx 3.31 \mu\text{g/kg}$ seawater, a value within 2% of the seawater value of Chen *et al.*³. The distilled water blank is therefore only 1.2% of the seawater value of a 10 ml sample. Because our accuracy was 2% despite a precision of better than 1%, we assume that our reported values are accurate to within 3% or better²². Reported E_h data were measured with a probe containing a platinum electrode which was regularly calibrated with Zobel solution. The probe was inserted into the sediment to a depth of 1 cm. There should have been no oxygen contamination at that depth because the cores were never exposed to air for more than a few seconds. Because a mixture of redox effects is measured, the results do not necessarily have any thermodynamic meaning other than indicating redox potentials on a relative scale. The operational use of E_h measurements is probably restricted to stagnant environments where oxygen has been replaced by other electron acceptors, except sulphate²³.

(GME, Madeira Abyssal Plain, Fault Area, $31^\circ 26.79' \text{ N}$, $24^\circ 48.75' \text{ W}$) and 60 (Southern Nares Abyssal Plain, silty, clayey turbidites, $23^\circ 34.49' \text{ N}$, $63^\circ 31.30' \text{ W}$) with a STACOR coring apparatus during the ESOPE cruise (June/July 1985). Three different methods were used to collect pore waters after sediment was placed in an inert N_2 -atmosphere at 4°C (ref. 7). (1) Centrifugation of sediment samples in 50 ml centrifuge tubes in a nitrogen-filled glove box in a cold van, in a Sorvall RC-2B refrigerated centrifuge at $10^4 g$ for 1 h, and subsequent $0.4 \mu\text{m}$ Nucleopore filtration of the pore water under N_2 . (2) Hydraulic (jack) squeezing of cold sediment, under nitrogen atmosphere, at 100 bar, through a Whatman 542 filter, and subsequent $0.45 \mu\text{m}$ Nucleopore filtration of the pore water under N_2 . (3) Nitrogen pressure squeezing in a nitrogen-filled glove box (modified from ref. 8), and subsequent filtration through a $0.2 \mu\text{m}$ cellulose acetate membrane filter.

A volume of 20–30 ml of pore water was collected from each interval extending 7–10 cm depth. Then 1–2 ml sub-samples from different intervals were combined and acidified to pH 1–2, with HCl, to obtain a composite 10 ml sample for uranium analysis by mass-spectrometric isotope-dilution methods. Each of these pore water samples had 6–12 months to equilibrate with O_2 .

Table 1 Uranium content of interstitial water samples of GME (core 24) and Nares (core 60) sediment

Sample number	Depth range (cm)	Collection method*	^{238}U concentration (ng g $^{-1}$)	E_h sediment† (mV)	NO_3 (μM)	pH (sediment)	Alkalinity (meq l $^{-1}$)	PO_4 (μM)	Si (μM)	Mn (μM)	Fe (μM)
<i>Core GME 24 (porosity 80–60%)</i>											
17722	18–106	2	2.10 ± 0.06	277	1.2	7.67	2.58	5.5	250	0.5	0.7
17734	310–408	1	0.63 ± 0.03	125	1.0	7.54	2.59	7.3	338	24	9.0
17746	608–707	1	0.57 ± 0.03	71	2.3	7.69	3.37	9.4	520	13	4.6
17758	908–1,005	1	0.55 ± 0.03	118	1.3	7.57	3.72	8.0	520	17	2.2
17770	1,206–1,315	1	0.13 ± 0.015	50	2.0	7.52	5.32	3.5	400	23	8.7
17785	1,583–1,676	1	(0.05 ± 0.01)	179	1.7	7.41	5.66	4.5	170	27	0.8
17802	2,008–2,102	2	0.92 ± 0.03	137	4.6	7.58	4.83	4.9	295	24	8.2
17814	2,301–2,400	2 (80%) 3 (20%)	$(2.12 \pm 0.07)^\ddagger$	64	5.3	7.52	4.60	6.7	507	16	1.3
17825	2,603–2,737	3	0.16 ± 0.015	41	2.3	7.54	4.83	8.2	618	15	2.1
17835	2,881–2,982	2	1.00 ± 0.03	78	3.4	7.50	4.76	6.6	540	15	7.0
<i>Core SNAP 60 (porosity 70–50%)</i>											
18100	15–144	1 (50%) 3 (50%)	0.85 ± 0.03	188	2.0	7.93	3.21	5.5	153	56	24
18107	176–303	1	1.93 ± 0.025	241	0.3	7.93	3.63	6.4	130	90	0.5
18113	343–508	2 (80%) 1 (20%)	1.21 ± 0.035	171	1.8	7.98	3.76	5.6	125	79	0.3
18119	541–777	2	0.55 ± 0.025	112	1.8	7.97	3.93	4.0	166	67	0.9
18127	814–1,004	2 (80%) 3 (20%)	0.38 ± 0.02	64	2.2	7.88	4.41	2.9	150	58	15
18134	1,037–1,216	3	0.13 ± 0.015	48	1.3	7.97	4.43	9.0	134	53	86
18140	1,246–1,404	3 (70%) 1 (30%)	0.17 ± 0.025	54	1.2	7.86	4.50	5.5	126	47	59
18146	1,439–1,641	1	0.49 ± 0.02	37	1.1	7.90	4.75	0.8	107	36	3.0
18153	1,672–1,872	1	0.49 ± 0.02	54	1.4	7.87	4.75	0.8	119	33	7

* Values in brackets signify fraction of volume collected by the method indicated. Method 2 tends to give highest uranium concentrations, method 3 tends to give lowest, and method 1 is intermediate.

† E_h was measured on board the ship by platinum electrodes punched into the sediments. The values reported here are depth averaged ($1\sigma \approx 20$ mV).

‡ This value has probably undergone air oxidation as the NO_3 concentration is too high.

Because of the rapid oxidation kinetics of U(IV) in the presence of O_2 (ref. 9), we can assume that regardless of the oxidation state of uranium in the sediment pore waters, it was analysed as U(VI).

The results of the pore water analysis of uranium and supporting geochemical parameters are shown in Table 1. In general, smooth pore water profiles of alkalinity, pH, Si, Mn and NH_4 (not shown) resulted, despite the fact that solid-phase profiles showed considerably larger down-core variations because of the turbiditic nature of the sediments.

All measured uranium concentrations are lower than the seawater uranium content, ranging from 2.1 to 0.85 p.p.b. near the surface of these cores to 0.05–0.5 p.p.b. at greater depth. The most reliable values can be considered to be those that had been extracted by method 3 (N_2 pressure squeezing). They average 0.15 ± 0.02 p.p.b. from both cores. The subtracted blank represents only about 20% of these values, making them still highly significant.

Some of the values of >0.5 p.p.b. at depth might have been caused by air oxidation of solid-phase U(IV) and subsequent release of U(VI) to pore waters. This is likely for extraction methods 1 and 2. When the results of the pore water analysis for nutrients and trace metals at GME are compared for the different methods, both methods 1 (centrifugation in presence of N_2) and 2 (hydraulic squeezing) gave Fe and NO_3 concentrations which indicated some air oxidation of Fe(II) and NH_3 (ref. 11).

In the zone where Fe(II) is mobilized [i.e. $E_h \leq 100$ mV] higher values of NO_3^- and uranium concentration are *a priori* suspect. For example, the presence of $5.3 \mu\text{M}$ NO_3^- , the highest value measured, in sample 17814, 2,301–2,400 cm depth of ESOPE core GME-24 makes the high uranium concentration of $2.12 \mu\text{g kg}^{-1}$ highly questionable. We therefore shall not consider this sample in any further discussion.

Below the first metre of sediment, the chemical conditions in the sediment pore waters are consistent with iron remobilization at $E_h \leq 150$ mV (ref. 12). At values of $E_h \sim 0$ –100 mV, $[\text{Fe}^{2+}] \approx 6 \mu\text{M}$ (maximum) was measured for core 24 and $89 \mu\text{M}$ for core 60, when only the best pore water collection methods (method 3) are considered. At measured alkalinity values between 2.5 and 6 meq l $^{-1}$ (possibly implying influence of SO_4^{2-} reduction at depth), PO_4 concentrations of $\sim 10 \mu\text{M}$ and pH of ~ 7.5 –8.0 (ref. 7), the uranium concentration could be expected to be a function of the carbonate and PO_4 -concentration¹, but [U] correlates with neither.

U(IV)-species is highly susceptible to oxidation in the presence of dissolved O_2 . Within 30 min, more than 90% of U(IV) oxidizes to U(VI)⁹. The presence of even small amount of O_2 could therefore increase the solubility of newly formed UO_2 . Indeed, it was observed^{4,13} that air contamination can increase the uranium concentrations in the pore waters by up to an order of magnitude. The most reliable pore water extraction method with respect to air contamination, method 3, produced pore water uranium concentrations below 1 m depth of only 0.15 ± 0.02 p.p.b. The lowest uranium concentration measured was 0.05 p.p.b. by method 1 (centrifugation) in pore waters from core 24 (GME). The solubility of UO_2 – U_4O_9 or USiO_4 , calculated from stability constants determined by Jordi *et al.*¹⁴, under the prevalent chemical conditions in the pore water, is, however, 2–3 orders of magnitude higher, making it likely that an adsorbed phase controls U(IV) solubility.

This very low value of 0.05 p.p.b. occurs in the region where a dissolved Si minimum (and an E_h maximum) occurs. It is thought that Si and possibly Fe are naturally being removed from solution here, possibly as FeSiO_3 or $\text{Fe}(\text{OH})_3$ with adsorbed H_3SiO_4^- (refs. 8, 15, 16). The same process responsible for the anomalies of Si and possibly Fe might also be responsible for the very low uranium concentration at that depth.

If we therefore exclude this uranium concentration and the uranium concentration with the highest down-core NO_3 concentration (indicating air oxidation of NH_3) from a correlation with other geochemical parameters, we find a significant ($P < 0.001$) linear correlation with the measured redox potential in the sediments (Fig. 1). As E_h approaches 0 mV, uranium concentration approaches a value of 0. Bottom water redox potentials, with the same measuring equipment, reached 500 mV. The least squares fit to the data points (see Fig. 1) would predict seawater values at 500 mV of 3.6 p.p.b., which is only 10% higher than the value of 3.24 p.p.b. (ref. 3). We therefore feel that the uranium concentrations in the pore waters somehow record the prevalent E_h conditions, but at thermodynamic control of uranium concentrations in pore waters, one would expect $\log [U]$ to be related to E_h . Even though Fig. 1 shows also a highly significant correlation of $\log [U]$ against E_h ($P \leq 0.001$), the correlation line does not have the expected Nernstian slope of $59/2 = 29.5$ mV. This might indicate that solubility of uranium is also related to other factors that co-vary with E_h .

Concentrations of $0.5\text{--}0.05 \mu\text{g kg}^{-1}$ of uranium, as measured in many of these pore water samples, are among the lowest concentrations reported for deep-sea sediments. Values of about 2.5 p.p.b. were reported for pore waters from the Pacific¹⁷. The lowest value reported for sulphate reducing conditions⁴ was 1.2 p.p.b., but most values were higher (up to 16 p.p.b.). Low uranium concentrations in these sediments coincided with the onset of Fe reduction, and higher values were hypothesized to arise from decomposition of organic matter and concomitant release of associated uranium in solution. The presence of authigenic uranium in settling organic matter was indentified by Anderson¹⁸ in the Atlantic and Pacific Oceans. Pressure artefacts in carbonate containing sediments could both decrease and increase pore water concentrations of uranium¹⁹. But the alkalinity difference between Kastenlot core and *in situ* pore water samples as measured by Toole *et al.*¹⁹ appears to be too large to be caused solely by pressure artefacts. Their *in situ* uranium concentration in oxic surface sediments are higher than seawater values, ranging from 3 to 8 p.p.b. They also report uranium concentrations for a box core from Nares Abyssal Plain sediments with 1.8–2.8 p.p.b., for 0–30 cm depth, somewhat higher than our value of 0.85 p.p.b. for the sample at 15–144 cm depth. But it is unlikely that the ESOPE cores sampled the top sediment^{7,20}.

As these pore water samples integrate uranium concentrations over ~1 m depth, the results are unsuited for any flux calculation. But they still allow an estimate of an upper limit for the mean diffusion path in the sediment: for core 24, solid-phase uranium concentrations vary from 1 to 4 p.p.m. (ref. 21) within a 1-m depth interval, with authigenic uranium concentrations ≤ 3 p.p.m. The highest apparent K_D -values (solid-phase to solution-phase concentration ratio) of authigenic uranium are then about $2 \times 10^4 \text{ cm}^3 \text{ g}^{-1}$, suggesting very limited U-mobility in the 8×10^5 years²¹ of accumulation of sediments in this core. For a mean diffusion path length, $\bar{x} \approx \sqrt{(D_{\text{eff}} t)}$ and $D_{\text{eff}} \approx \phi^2 D_m / K_D$ (where ϕ represents average porosity ≈ 0.5 , D_m represents molecular diffusion coefficient of U(IV), taken as $10^{-6} \text{ cm}^2 \text{ s}^{-1}$), $\bar{x}$ is only 30 cm.

The pore water results are therefore in agreement with the solid-phase data^{5,21}, which suggest little uranium mobility in these turbiditic sediments. Uranium concentration maxima in the lowermost sections of this core must have persisted for about 7.5×10^5 years (ref. 21).

These results were obtained from the international ESOPE expedition, the Marion Dufresne cruise MD45, coordinated by the NEA/Seabed working Group. We acknowledge the help and support given by the ESOPE scientific party and the ESOPE funding agencies, and the support given to this project by H. Gaggeler (EIR). We thank R. F. Anderson for helpful suggestions and for reviewing the manuscript, and G. J. de Lange for relating the redox potentials to us.

Received 6 May; accepted 6 October 1987.

- Langmuir D. *Geochim. cosmochim. Acta* **42**, 547–569 (1978).
- Cochran, J. K. *Uranium Series Disequilibrium: Applications to Environmental Problems* (eds Ivanovich, M. & Harmon, R. S.) 384–430 (Clarendon, Oxford, 1982).
- Chen, J. H., Edwards, R. L. & Wasserburg, G. J. *Earth planet. Science Lett.* **80**, 241–251 (1986).
- Cochran, J. K., Carey, A. E., Sholkovitz, E. R. & Suprenant, L. D. *Geochim. cosmochim. Acta* **50**, 669–680 (1986).
- Colley, S., Thomson, J., Wilson, T. R. S. & Higgs, N. C. *Geochim. cosmochim. Acta* **48**, 1223–1235 (1984).
- Colley, S. & Thomson, J. *Geochim. cosmochim. Acta* **49**, 2339–2348 (1985).
- Cranston, R. *et al.* in *Geoscience Investigations of two North Atlantic Abyssal Plains—ESOPE International Expedition* (eds Schuttenhelm, R. T. E. *et al.*) (CEC-Joint Research Center, JRC Report EUR, in the press).
- Reeburgh, W. S. *Limnol. Oceanogr.* **12**, 163–165 (1967).
- Anderson, R. F. *Nucl. Instrum. Meth. Phys. Res.* **223**, 213–217 (1984).
- Keil, R. Z. *analyst. Chem.* **297**, 384–387 (1979).
- De Lange *et al.* in *Geoscience Investigations of two North Atlantic Abyssal Plains—ESOPE International Expedition* (ed. Schuttenhelm, R. T. E. *et al.*) (CEC-Joint Research Center, JRC Report EUR, in the press).
- De Lange, G. J. *Geochim. cosmochim. Acta* **50**, 2543–2561 (1986).
- Anderson, R. F. *Eos* **67**, 1069 (1986).
- Bruno, J., Forsyth, R. S. & Werme, L. O. *Scientific Basis for Nuclear Waste Management VII* (eds Jantzen, C. M., Stone, J. & Ewing, R. C.) 413–421 (Materials Research Society, Pittsburgh, Pennsylvania, 1985).
- Hydes, D. J., Cranston, R. E. & De Lange, G. in *Geoscience Investigations of two North Atlantic Abyssal Plains—ESOPE International Expedition* (eds Schuttenhelm, R. T. E. *et al.*) (CEC-Joint Research Center, JRC Report EUR, in the press).
- De Lange, G. J. & Rispens *Mar. Geol.* **73**, 85–97 (1986).
- Cochran, J. K. & Krishnaswami *Am. J. Sci.* **280**, 849–889 (1980).
- Anderson, R. F. *Geochim. cosmochim. Acta* **46**, 1293–1299 (1982).
- Toole, J., Thomson, J., Wilson, T. R. S. & Baxter, M. S. *Nature* **308**, 263–266 (1984).
- Auffret, G. A. *et al.* in *Geoscience Investigations of two North Atlantic Abyssal Plains—ESOPE International Expedition* (eds Schuttenhelm, R. T. E. *et al.*) (CEC-Joint Research Center, JRC Report EUR, in the press).
- Colley, S. in *Geoscience Investigations of two North Atlantic Abyssal Plains—ESOPE International Expedition* (eds Schuttenhelm, R. T. E. *et al.*) (CEC-Joint Research Center, JRC Report EUR, in the press).
- Santschi, P. H., Bajo, C., Mantovani, M., Orcinolo, D., Cranston, R. & Bruno, J. in *Geoscience Investigations of two North Atlantic Abyssal Plains—ESOPE International Expedition* (eds Schuttenhelm, R. T. E. *et al.*) (CEC-Joint Research Center, JRC Report EUR, in the press).
- Whitfield, M. *Limnol. Oceanogr.* **19**, 857–865 (1974).

Magma–cumulate mixing identified by U–Th disequilibrium dating

D. M. Pyle,*† M. Ivanovich* & R. S. J. Sparks†

* Uranium Series Disequilibrium Section, Nuclear Physics Division H7, AERE Harwell, Oxon OX11 0RA, UK
† Department of Earth Sciences, Downing Street, Cambridge CB2 3EQ, UK

Magma mixing is believed to be important in magmagenesis¹, particularly in volcanic arcs², where its influence is revealed by heterogeneous phenocryst populations³ and quenched inclusions⁴. Nevertheless isotope studies, particularly of uranium-series disequilibria, have concentrated on the analysis of bulk rock samples^{5–9}; comprehensive U–Th mineral-separate data from arc volcanics are scarce^{10–14}. Here we present the first U–Th isotopic evidence for mixing between a crystal mush and a magma in an andesite from the island of Santorini, in the Aegean arc. The lava contains crystal populations from two sources of distinct thorium isotope composition: one from a basic cumulate; the other phenocrysts from a dacite magma. The crystal populations have statistically indistinguishable crystallization ages of 79_{-12}^{+14} kyr and 93_{-22}^{+29} kyr, and are thought to have physically intermingled just before eruption. The crystals did not re-equilibrate and hence retain the isotopic compositions and ages of the parental materials.

In an evolving magma chamber, the $^{230}\text{Th}/^{232}\text{Th}$ activity ratio will vary with time as the short-lived ^{230}Th strives to reattain radioactive equilibrium with its parent ^{234}U . Arrival of new melt batches will also vary the $^{230}\text{Th}/^{232}\text{Th}$ activity ratio of the system. Thus, during growth, phenocrysts could become isotopically zoned with respect to ($^{230}\text{Th}/^{232}\text{Th}$). Consequently, the relative amounts of zoned phenocryst cores and rims will determine the isotopic composition of any mineral separate. Unlike the other major isotopic systems, two magma batches from the same source region which have followed identical evolutionary histories will not have similar Th isotope characteristics, unless

Table 1 Mineral separates

Sample	Harwell code	Density	Description
Bulk rock	4718		<90 mesh fraction from whole rock (more crystal-rich than lava)
Plagioclase	4719	$\rho > 2.70$	Calcic (bytownite) cores; ~70% of total plagioclase separated
Olivine	4721	$\rho < 3.6$	Homogeneous crystals, with minor apatite and clinopyroxene
Clinopyroxene	4722	$\rho > 3.35$	Iron-rich augite cores
Magnetite	4723	$\rho > 4.0$	Total population (magnetic)
Zircon	4725	$\rho > 4.0$	Total population (non-magnetic)

their extraction times are coincidentally equal¹⁰. Mixing of chemically identical melts, producing petrographically equilibrated products, could yield isotopically disequibrated samples with respect to ($^{230}\text{Th}/^{232}\text{Th}$).

An example of Th isotopic heterogeneity was reported by Capaldi and Pece¹³, who analysed mineral separates from several historic lavas. They found ($^{230}\text{Th}/^{232}\text{Th}$) disequilibrium between phases in each sample. Isotopic heterogeneity was confirmed with lead isotope data¹⁵. They concluded that the melt phase was isotopically inhomogeneous and that the uranium series method could not be used for dating volcanic materials. Their results and conclusions were disputed by Hemond and Condomines¹⁶, who found no evidence for isotopic heterogeneity in the same samples, suggesting that these lavas had formed within a closed system.

Lava S8623 is a typical flow-banded, crystal-rich, weakly vesicular basaltic andesite from the latest products of the Mikro Profitis Ilias shield in northern Santorini, Greece (Fig. 1). The stratigraphy and geochemistry of this shield have already been documented¹⁷⁻¹⁹. It corresponds to the lava P4.6 of Huijsmans¹⁷ and lies between the 100 ± 30 kyr Lower Pumice series (Bu) (refs 20, 21) and the >37 kyr Middle Tuff sequence²².

Petrological evidence for disequilibrium in P4.6 was presented by Huijsmans¹⁷, who found two plagioclase feldspar populations: a normally zoned calcic population ($\text{An}_{82}\text{-An}_{71}$), and a relatively sodic population ($\text{An}_{69}\text{-An}_{55}$) showing resorption textures. Three fields of clinopyroxene compositions were defined, corresponding to populations of Mg-rich cores, Fe-rich cores and intermediate rims. Embayed olivines ($\text{Fo}_{81}\text{-Fo}_{69}$) and magnetites were also reported.

Pure mineral separates of the crystal phases were prepared using standard methods (Table 1). Following total dissolution and spiking with ^{236}U and ^{229}Th , uranium and thorium were separated and purified on anion exchange columns, and the electroplated sources analysed by alpha spectrometry. Results are presented in Table 2 and plotted on a $^{230}\text{Th}/^{232}\text{Th}$ against $^{238}\text{U}/^{232}\text{Th}$ isochron diagram (Fig. 2).

The six samples appear to define two sub-parallel arrays (Fig. 2). The alternative interpretation, of several arrays of mineral pairs, can be discounted as physically unreasonable (for example, Mg-olivine and zircon cannot crystallize together) or

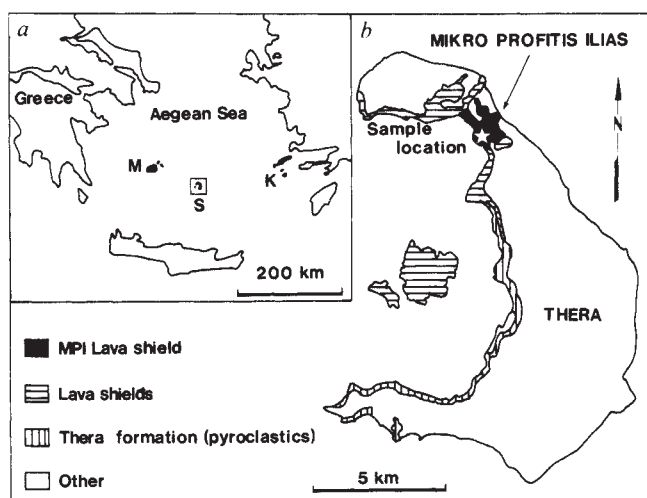


Fig. 1 a, Location of recent volcanism in the Aegean. M, Milos; S, Santorini; K, Kos, Nisiros and Yali. b, Location of Mikro Profitis Ilias Lava Shield. Adapted after Huijsmans¹⁷ and references therein. Thera Formation refers to the pyroclastic products of the Thera volcano²¹.

otherwise give unrealistic 'ages'. The high ($^{230}\text{Th}/^{232}\text{Th}$) 'olivine'-bytownite-magnetite component represents a plausible crystallizing assemblage from a basaltic liquid with an initial ($^{230}\text{Th}/^{232}\text{Th}$) ratio of $1.33^{+0.23}_{-0.13}$. The high thorium concentration of the 'olivine' is due to the presence of small apatite inclusions within the separate. The low ($^{230}\text{Th}/^{232}\text{Th}$) assemblage ('bulk rock', iron-rich augite, zircon) was derived from a more evolved liquid, with an initial ($^{230}\text{Th}/^{232}\text{Th}$) ratio of 0.88 ± 0.03 . Trace element data^{17,19} from Santorini show depletion of Zr, Hf and heavy rare earth elements (otherwise incompatible elements which fractionate strongly into zircon) after 65-68% SiO_2 , suggesting onset of zircon crystallization in dacitic magmas. The coexisting clinopyroxene ($\text{Wo}_{41}\text{En}_{44-42}\text{Fs}_{16-18}$) and plagioclase compositions are consistent with reported phenocryst compositions of Santorini dacites to rhyodacites¹⁷. The lack of modal orthopyroxene in S8623 can only be reconciled with its expected crystallization from andesitic, or more evolved, liquids if the magma had $\text{SiO}_2 < 68\text{-}70\%$ (orthopyroxene becomes dominant only after this) or was crystal poor.

A simple mass-balance calculation can constrain the extent of mixing. The high ($^{230}\text{Th}/^{232}\text{Th}$) component has not shifted the 'bulk rock' thorium composition detectably from the low ($^{230}\text{Th}/^{232}\text{Th}$) array. Our unpublished data suggest that the concentration ratio $\text{Th}_{(\text{dacite})}/\text{Th}_{(\text{basalt})}$ is <4 . Using this and the silica content as constraints, the basaltic liquid input (SiO_2 50%) must have been less than ~24 weight % (of the total magma erupted) to change the bulk thorium composition by less than 5% (the maximum undetectable shift). This satisfies the SiO_2 balance, producing a basaltic andesite (SiO_2 55%), after allowing for the crystal content of S8623 (ref. 17) (olivine 4%, plagioclase 23%, clinopyroxene 7%, magnetite $<1\%$, bulk SiO_2 47%). As $>70\%$ of the plagioclase and the bulk of the magnetite came from the basalt, this component had a minimum crystal content of 48%

Table 2 Isotope data from Santorini

Sample	Harwell code	Concentrations (p.p.m.)		$(^{238}\text{U}/^{232}\text{Th})$	Activity ratios $(^{230}\text{Th}/^{232}\text{Th})$	$(^{232}\text{U}/^{238}\text{U})$
		U	Th			
Bulk rock	4718	$1.013 \pm 0.049^*$	4.12 ± 0.12	0.753 ± 0.042	0.811 ± 0.029	1.13 ± 0.07
Plagioclase	4719	0.106 ± 0.005	0.392 ± 0.017	0.82 ± 0.06	1.06 ± 0.07	1.08 ± 0.06
Olivine	4721	0.715 ± 0.024	5.77 ± 0.18	0.379 ± 0.018	0.781 ± 0.028	0.996 ± 0.040
Clinopyroxene	4722	0.490 ± 0.018	1.48 ± 0.07	1.02 ± 0.06	0.96 ± 0.06	0.978 ± 0.043
Magnetite	4723	1.76 ± 0.07	4.96 ± 0.21	1.09 ± 0.06	1.18 ± 0.06	1.09 ± 0.05
Zircon	4725	$2,360 \pm 60$	$3,710 \pm 120$	1.95 ± 0.08	1.431 ± 0.045	0.983 ± 0.022

* One sigma errors from counting statistics only.

when mixing occurred. The contaminated dacite had a maximum crystal content of 23%. The high ($^{230}\text{Th}/^{232}\text{Th}$) component is thus interpreted as a crystal-rich basaltic mush which was disrupted by, and efficiently mixed with a dacite magma before eruption.

The two populations crystallized at 79^{+14}_{-12} kyr and 93^{+29}_{-22} kyr (Fig. 2) from liquids of distinct ($^{230}\text{Th}/^{232}\text{Th}$) compositions. These ages are consistent with previously reported dates^{20,22} and are indistinguishable within 1σ , implying that the basaltic mush formed less than ~ 14 kyr before the intrusion of the dacite magma. The dacite crystallization age (79^{+14}_{-12} kyr) represents a maximum eruption age for S8623. The difference between the ($^{230}\text{Th}/^{232}\text{Th}$) activity ratios of the parental magmas is similar to the total reported variation from zero-age lavas^{6,23} and may reflect: (1) melting of sources of distinct ($^{230}\text{Th}/^{232}\text{Th}$) composition, (2) melting of the same source but with different melt extraction times, and (3) influence of crustal (low ($^{230}\text{Th}/^{232}\text{Th}$)) material in the generation of the more evolved magma. These alternatives could be distinguished by lead or strontium isotope analysis. There is no indication of the uranium enrichment believed to be characteristic of arc magmas^{5,10,23}, and all samples have $^{234}\text{U}/^{238}\text{U} \sim 1.0$ within 2σ . Mixing clearly was important in the generation of this lava, but as a mingling of a crystal mush and a dacitic magma, rather than of two magmas.

In conclusion, petrographic disequilibrium is mirrored by ($^{230}\text{Th}/^{232}\text{Th}$) heterogeneity. This is anticipated to be common in calc-alkaline settings, and could explain variability in the data presented by Allegre and Condomines¹⁰. Mineral separates may be used for U-Th dating only when fully equilibrated coexisting phases, or parts of phases (such as rims), are used. Other separates may be used to trace the magmatic evolution of each crystal phase and the environment from which they crystallized, for example by analysing tightly controlled density (hence compositional) plagioclase separates. The full power of mineral separate data will only be realized when meaningful samples are used.

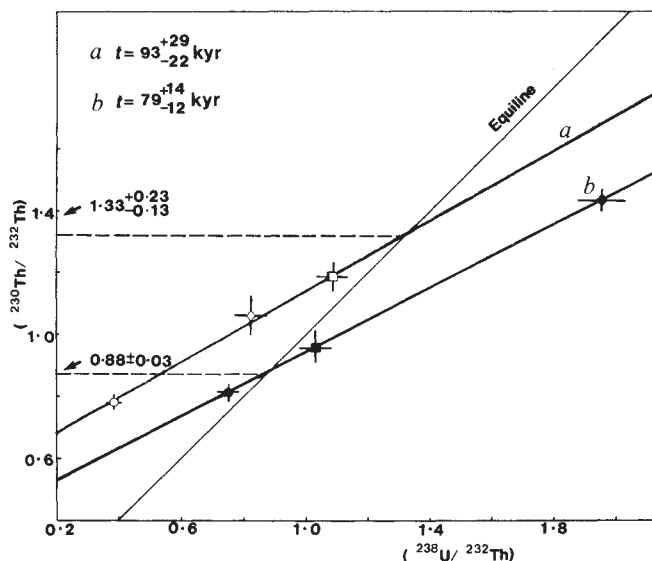


Fig. 2 U-Th isochron diagram for S8623. The ages were obtained from the isochron slopes, which were calculated using a least-squares fitting program for variables with correlated errors²³. The slope m is related to the age t by $m = (1 - \exp[-\lambda t])$, where λ is the decay constant for ^{230}Th ($9.22 \times 10^{-3} \text{ kyr}^{-1}$). The equiline is the locus of fully equilibrated samples, with $^{230}\text{Th}/^{238}\text{U} = 1$. The initial $^{230}\text{Th}/^{232}\text{Th}$ ratio of the system is given by the intercept of the isochron with the equiline, or by $(^{230}\text{Th}/^{232}\text{Th})_i = (^{230}\text{Th}/^{232}\text{Th})_0 \cdot \exp[-\lambda t]$ (ref. 11), where $(^{230}\text{Th}/^{232}\text{Th})_i$ is the intercept on the $^{230}\text{Th}/^{232}\text{Th}$ axis. Filled symbols, (◆) zircon, (●) augite, (●) total rock; open symbols, (◇) calcic plagioclase, (□) magnetite, (○) olivine.

DMP acknowledges the director of IGME for permission to carry out field work in Greece and the NERC for a studentship. We thank S. Hasler, M. Haslop and M. Wilkins for technical support, Dr G. Longworth for the least-squares program and R. K. O'Nions and two anonymous reviewers for useful comments on the manuscript. R.S.J.S. acknowledges the support of the B.P. Venture Research Unit. Radiochemical work was carried out in the Uranium Series Disequilibrium Section, Harwell Laboratories. Work described in this report was undertaken as part of the Underlying Research Programme of the UKAEA.

Received 12 August; accepted 6 November 1987.

- Anderson, A. T. *J. Volcan. geotherm. Res.* **1**, 1-33 (1976).
- Eichelberger, J. C. *Nature* **275**, 21-27 (1978).
- Sakuyama, M. *J. Petrol.* **22**, 553-583 (1981).
- Eichelberger, J. C. *Nature* **288**, 446-450 (1980).
- Newman, S., MacDougall, J. D. & Finkel, R. C. *Nature* **308**, 268-270 (1984).
- Newman, S., MacDougall, J. D. & Finkel, R. C. *Contr. Miner. Petrol.* **93**, 195-206 (1986).
- Bennett, J. T., Krishnaswami, S., Turekian, K. K., Melson, W. G. & Hopson, C. A. *Earth planet. Sci. Lett.* **60**, 61-69 (1982).
- Capaldi, G. et al. *Bull. volcan.* **41**, 259-285 (1978).
- Capaldi, G., Cortini, M. & Pece, R. *Isot. geosci.* **1**, 39-55 (1983).
- Allegre, C. J. & Condomines, M. *Earth planet. Sci. Lett.* **28**, 395-406 (1976).
- Condomines, M. & Allegre, C. J. *Nature* **288**, 354-357 (1980).
- Condomines, M., Tanguy, J. C., Kieffer, G. & Allegre, C. J. *Geochim. cosmochim. Acta* **46**, 1397-1416 (1982).
- Capaldi, G. & Pece, R. *J. volcan. geotherm. Res.* **11**, 367-372 (1981).
- Capaldi, G., Cortini, M. & Pece, R. *J. volcan. geotherm. Res.* **14**, 247-260 (1982).
- Cortini, M. & van Calsteren, P. W. C. *Nature* **314**, 343-345 (1985).
- Hemond, C. & Condomines, M. *J. volcan. geotherm. Res.* **26**, 365-368 (1985).
- Huijsmans, J. P. P. *Geol. Ultraetina* **41**, 1-316 (1985).
- Nicholls, I. A. *J. Petrol.* **12**, 67-119 (1971).
- Mann, A. C. *Contr. Miner. Petrol.* **84**, 43-57 (1983).
- Seward, D., Wagner, G. A. & Pichler, H. in *Thera and the Aegean World Vol. II* (ed. Doumas, C.) 101-108 (1980).
- Pichler, H. & Kussmaul, S. N. *Jb. Miner. Abh.* **116**, 268-307 (1972).
- Friedrich, W. L., Pichler, H. & Kussmaul, S. *Bull. geol. Soc. Denmark* **26**, 27-39 (1977).
- Allegre, C. J. & Condomines, M. *Nature* **299**, 21-24 (1982).
- York, D. *Earth planet. Sci. Lett.* **5**, 320-324 (1969).

Melting of vapour-absent tonalite at 10 kbar to simulate dehydration-melting in the deep crust

Michael J. Rutter & Peter J. Wyllie

Division of Geological and Planetary Sciences 170-25,
California Institute of Technology, Pasadena, California 91125, USA

Granitoids may be formed by differentiation of mantle-derived basalts, by partial fusion of crustal rocks, or by interaction of melts from these two sources. Except in special circumstances, the amount of pore fluid in the deep continental crust is likely to be so small that partial melting of rocks will take place under vapour-absent conditions, with water becoming available through dehydration-melting reactions. Here we present the results of experiments in which garnet tonalite is melted at 10 kbar in the absence of vapour. Between 825 and 900 °C the biotite melting reaction produces 20% melt; by 1,000 °C hornblende has melted, yielding 35% melt. As a melt fraction of 30-50% is required before a granodioritic melt can be segregated from its source rock²¹ our results imply that, whereas migmatites may be generated at typical high-grade metamorphic temperatures of 825 °C, the segregation of large-volume granitoid magmas from orthogneisses requires temperatures greater than 950 °C. In the absence of large amounts of externally-derived aqueous fluid, underplating or emplacement of hot basalt into the crust is required to raise the temperature sufficiently for magma segregation.

Potential crustal source rocks include granulites, amphibolites, tonalitic gneisses and metamorphosed sedimentary rocks. The relatively high K/Ca implied by Ca isotope systematics for the source of Sierra Nevada granitoids¹ suggests that tonalitic-granodioritic gneisses may be abundant components of deep crustal terrains. It has been shown that mafic

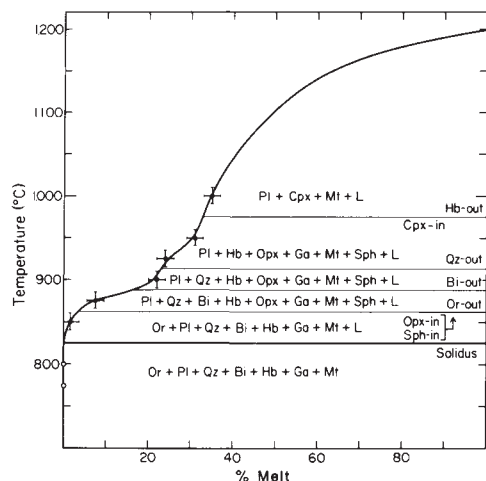


Fig. 1 Diagram showing the effect of temperature on the melting interval for tonalite 101 at 10 kbar. The liquidus temperature is taken from Stern *et al.*²². Open circles represent experiments conducted for one-day duration; closed circles experiments of 4- to 5-d duration. Error bars based $\pm 10^\circ\text{C}$ for temperature measurements in piston cylinder apparatus and $\pm 2\%$ error in the mode estimates.

garnet-bearing tonalites from the southernmost Sierra Nevada batholith equilibrated at 8 kbar (D. Sams and J. Saleeby, personal communication), corresponding to a depth of approximately 25 km.

There have been few experimental studies dealing with crustal magmas and vapour-absent melting regions for rocks containing muscovite, biotite or amphibole²⁻⁸ and, of these, only Brown and Fyfe² and Huang and Wyllie^{4,5} provided vapour-absent solidus curves. Phase diagrams for evaluation of dehydration-melting of rocks have been based on extrapolation from these experiments, phase equilibrium constraints and thermodynamic calculations⁹⁻¹². Several thermodynamic treatments of dehydration-melting reactions at reduced water activity have been published¹³⁻¹⁷. The most detailed analysis of melting in vapour-absent biotite and hornblende gneisses is that of Burnham¹².

Following extensive studies of gabbro-tonalite-granite- H_2O (ref. 8) we have melted these rocks under vapour-absent conditions at 10–15 kbar. We present here the results of melting a tonalite¹⁸ at 10 kbar, with 0.8 wt % H_2O stored in biotite (12.5 modal %) and hornblende (9 modal %). The rock did not initially contain garnet. Experiments were conducted in gold capsules in piston-cylinder apparatus for durations of 1 and 4–5 days, with negligible Fe-loss to the capsule. One-day runs were obviously not equilibrated, but reaction was well advanced after 4 days. The experiments were not reversed, but measured systematic changes in mineral and melt compositions indicate a close approach to equilibrium. Polished charges were examined using scanning electron microscopy, and analysed using an electron microprobe. Quenched melt was easily identified in melted runs as a low-contrast, morphologically conspicuous interstitial phase. The percentage of glass in each run was determined by point-counting SEM photomicrographs.

The results in Fig. 1 illustrate the progressive melting of garnet tonalite containing the minerals alkali feldspar, plagioclase, quartz, biotite, hornblende, garnet and magnetite. Garnet grew subsolidus at 10 kbar in addition to the minerals originally present in the tonalite. The solidus reaction at $825 \pm 25^\circ\text{C}$ was the initial breakdown of biotite releasing water for solution in H_2O -undersaturated liquid; biotite exists through a melting interval of about 70°C . The appearance of orthopyroxene is consistent with the solidus reaction reported by Peterson and Newton¹⁹ in $\text{K}_2\text{O-MgO-Al}_2\text{O}_3\text{-SiO}_2\text{-H}_2\text{O}$ at 10 kbar, although

rutile is also produced during biotite breakdown in the TiO_2 -bearing natural tonalite. The hornblende reaction, which releases more water for the melt, is completed between 950 and $1,000^\circ\text{C}$ with the generation of clinopyroxene. The other minerals melt as shown in Fig. 1 and, at temperatures above the hornblende-out boundary, clinopyroxene is the sole ferromagnesian mineral.

Figure 1 shows that the percentage of melted rock increases in a step-like fashion between 850 and 900°C where biotite reacts out, reaching 22% at 900°C . The data are consistent with a second smaller step between 925 and 975°C coinciding with where hornblende melts, although a smooth curve could equally be drawn through the error bars. At $1,000^\circ\text{C}$ the rock is 35% molten, and the liquid contains approximately 2.3 wt % H_2O . The sizes of these steps would vary as a function of modal percentage of biotite and hornblende¹².

There is no doubt that migmatites can be generated by *in situ* dehydration-melting of amphibolites and tonalitic gneisses, but there are questions about the conditions under which such melts could become segregated to provide magma for I-type granitoid intrusions²⁰. A partly molten rock can be mobilized to become a magma when the critical melt fraction (CMF) is reached, and it is argued that extraction of silicate melts can occur only when the CMF is exceeded. In a recent review, Wickham²¹ concluded that the CMF in rocks containing granitic melt was between 30 and 50%. Figure 1 shows that this condition during melting of tonalite is not achieved during dehydration-melting involving biotite. Between 30 and 35% liquid is generated during melting of hornblende; therefore a H_2O -undersaturated granodiorite magma (with about 2.3 wt % H_2O) could be segregated from a source at the base of the continental crust through dehydration-melting only at temperatures between 950 and $1,000^\circ\text{C}$. In this temperature range, the entire partly melted tonalite mass may begin to behave as a magma, capable of intrusion to a higher level.

The effect of aqueous pore fluid in reasonable quantities would be to introduce some melt at lower temperatures, but the conclusions would remain unchanged. Similar temperatures would be required for melt extraction from dehydration-melting of amphibolite^{11,12}, but pelitic gneisses with high percentages of biotite may yield sufficient melt at lower temperatures in the range 850 to 900°C (J. Holloway, personal communication). The question of the possible effects of deformation on the separation of viscous granitic melts from crustal rocks has not been experimentally investigated.

Hot basalt underplating or emplacement into the crust is probably a prerequisite to raising the temperature to that required for magma segregation from gneisses of igneous origin. This introduces the prospect of enhancing the amount of I-type granitoid melt at deep levels by interaction of the basaltic and granodioritic melts in ways discussed by Burnham¹².

Received 20 July; accepted 26 October 1987.

- Marshall, B. D. & De Paolo, D. J. *Eos* **68**, 465 (1987).
- Brown, G. C. & Fyfe, W. S. *Contrib. Miner. Petrol.* **28**, 310–318 (1970).
- Robertson, J. K. & Wyllie, P. J. *J. Geol.* **79**, 549–571 (1971).
- Huang, W. L. & Wyllie, P. J. *Contrib. Miner. Petrol.* **42**, 1–14 (1973).
- Huang, W. L. & Wyllie, P. J. *J. geophys. Res.* **86**, 1015–1079 (1981).
- Maaløe, S. & Wyllie, P. J. *Contrib. Miner. Petrol.* **52**, 175–191 (1975).
- Naney, M. T. *Am. J. Sci.* **238**, 993–1033 (1983).
- Huang, W. L. & Wyllie, P. J. *Am. Miner.* **71**, 301–316 (1986).
- Robertson, J. K. & Wyllie, P. J. *Am. J. Sci.* **271**, 252–277 (1971).
- Wyllie, P. J. *J. geophys. Res.* **76**, 1328–1338 (1971).
- Wyllie, P. J. *Tectonophysics* **13**, 41–71 (1977).
- Burnham, W. C. in *Geochemistry of Hydrothermal Ore Deposits* (ed. Barnes, H. L.) 71–136 (Holt, Rinehart and Winston, 1979).
- Thompson, A. B. *Am. J. Sci.* **282**, 1567–1595 (1982).
- Powell, R. J. *geol. Soc.* **140**, 629–633 (1983).
- Grant, J. A. in *Migmatites* (ed. Ashworth, J. R.) 86–144 (Blackie, Glasgow, 1985).
- Ellis, D. J. & Thompson, A. B. *J. Petrol.* **27**(1), 91–121 (1986).
- Waters, D. J. *J. metamorph. Petrol.* (in the press).
- Piwoński, A. J. *J. Geol.* **76**, 548–570 (1987).
- Peterson, J. W. & Newton, R. C. *Eos* **68**, 451 (1987).
- White, A. J. R. & Chappell, B. W. *Tectonophysics* **43**, 7–22 (1977).
- Wickham, S. M. *J. geol. Soc.* **144**(2), 281–299 (1987).
- Stern, C. R., Huang, W. L. & Wyllie, P. J. *Earth planet. Sci. Lett.* **28**, 189–196 (1975).

A nitrate-dependent *Synechococcus* bloom in surface Sargasso Sea water

Hilary E. Glover*, Barbara B. Prézélin†, Lisa Campbell‡, Michael Wyman§ & Christopher Garside*

* Bigelow Laboratory for Ocean Sciences, McKown Point, West Boothbay Harbor, Maine 04575, USA

† Department of Biological Sciences, Marine Science Institute, University of California, Santa Barbara, California 93106, USA

‡ Department of Oceanography, University of Hawaii, Honolulu, Hawaii 96822, USA

§ Department of Biological Sciences, University of Warwick, Coventry CV4 7AZ, UK

Considerable debate exists concerning the magnitude of oceanic primary production, its rate of transfer to other trophic levels and turnover times of carbon and nitrogen¹⁻⁵. In nitrogen-limited ocean systems, episodic increases in nitrate concentrations can support a significant fraction of annual phytoplankton production⁵. Yet little information is available regarding the distribution of nitrate in seasonally stratified oceanic surface waters, because concentrations are below the 0.03 μM detection limit of colorimetric methods⁶. We present the first evidence that high surface productivity in stratified Sargasso Sea water was supported by nanomolar changes in nitrate concentrations. This change was stoichiometrically consistent with the subsequent cellular production of a cyanobacterial (*Synechococcus*) bloom. Initially, cellular phycoerythrin and chlorophyll pigments increased, after which growth was enhanced to near maximum rates, and grazing was closely coupled to production. These observations suggest that *Synechococcus* occupies an important trophic position in the transfer of new nitrogen into the oceanic food web.

We report the use of two recently developed techniques to measure: (1) nitrate concentrations at the nanomolar level by chemiluminescence⁷; and (2) phycoerythrin pigment concentrations using uncoupled *in vivo* fluorescence⁸, in Sargasso-Sea surface waters. Together with observations made using established methods^{9,10}, these novel data provide new insight into the relationship between nutrient availability and phytoplankton distribution and production in the oligotrophic ocean. Data were collected between 20 and 28 July 1986, on research vessel (RV) *Endeavor* cruise EN146, at a station 140 km east of Bermuda (32°42'N, 63°01'W). Chlorophyll *a* concentrations, *in situ* rates of ¹⁴C-based primary production, and epifluorescence-microscopic enumeration of eukaryotic chlorophyll-fluorescing algae and prokaryotic phycoerythrin-fluorescing cyanobacteria, *Synechococcus* species were determined in four size fractions: 0.2–0.6 μm , 0.6–1 μm , 1–5 μm and >5 μm . *Synechococcus* accounted for 98 to 100% of all autofluorescent cells only in the 0.6–1.0 μm fraction. Cell specific *Synechococcus* properties from 0.6–1.0 μm data were multiplied by *Synechococcus* cell numbers in other size fractions to estimate *Synechococcus* properties for those fractions.

Between 20 and 25 July, a decrease in nitrate concentrations and a phytoplankton bloom occurred in surface waters (0–25 m). Cell numbers, pigment concentrations and measurements of primary productivity on size fractions of the total phytoplankton community are consistent with a transient bloom of *Synechococcus*. Figure 1 presents an overview of the temporal sequence of phytoplankton properties describing the bloom in surface waters (2–5 m).

On 21 July, there was an increase in chlorophyll concentrations in the 0.6–1.0 μm size range (Fig. 1b) which was dominated by *Synechococcus*. Concentrations in the principally algal >5.0 μm fraction and the mixed 1.0–5.0 μm assemblage showed relatively little change throughout the period. Chlorophyll and phycoerythrin per *Synechococcus* cell increased on 21 July (Fig. 1c and d), followed by an increase in *Synechococcus* abundance

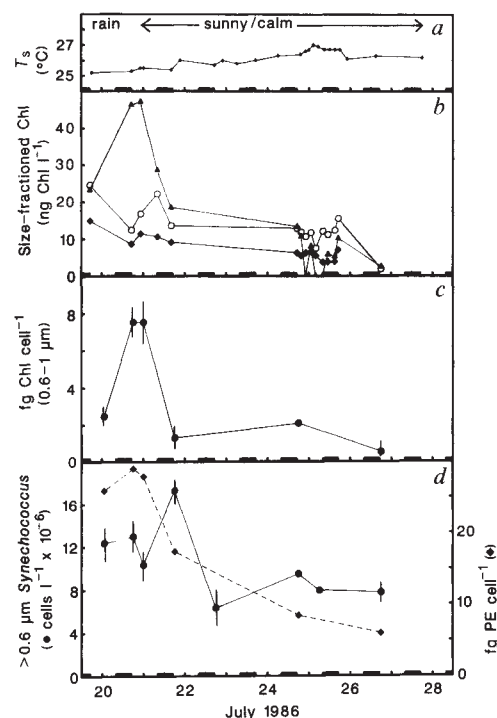


Fig. 1 Temporal changes in surface waters (2–5 m) between 20–28 July 1986 in the northern Sargasso Sea, at 32°42' N, 63°01' W. Graph a shows sea-surface temperature. Graph b shows chlorophyll concentrations⁹ in 0.6–1 μm ($\blacktriangle$), 1–5 μm ($\circ$) and >5 μm ($\blacklozenge$) communities. Graph c shows chlorophyll content in 0.6–1 μm *Synechococcus* cells ($\bullet$). Graph d shows abundance ($\bullet$)⁹ and cellular phycoerythrin content ($\blacklozenge$)⁸, of the >0.6 μm *Synechococcus* population.

(Fig. 1d). As cell numbers increased, cellular pigment concentrations declined rapidly but not at equal rates. Similar changes occurred at the same time at other depths in the upper 25 m and a quantitative description of the bloom can be presented in terms of 25-m integrated changes.

On 25 July, the nitrate profile (Fig. 2a) was remarkably similar to the only previous observation in the Sargasso Sea using this technique¹¹. On 20 July, however, concentrations in the upper 25 m (21–27 nM NO_3^-) were higher than those at any depth above the nitracline, and an order of magnitude higher than previous measurements¹¹. Between 20 and 25 July, 417 $\mu\text{mol NO}_3^- \text{m}^{-2}$ was removed from the upper 25 m, coincident with a bloom of *Synechococcus* over the same depth range (Table 1).

Integrated chlorophyll increased between 20 and 21 July and decreased to 50% of its original concentration by 25 July. Virtually all of the change between the first and second day could be accounted for by an 85% increase in *Synechococcus* chlorophyll. Thereafter, concentrations in both the total phytoplankton assemblage and the pure *Synechococcus* 0.6–1.0 μm fraction declined about 2.5-fold by 25 July. By contrast, there was a fivefold decline in phycoerythrin concentrations over the five days and phycoerythrin to chlorophyll ratios in *Synechococcus* (all fractions) declined from 13.8 to no more than 6.5. Changes in this pigment ratio were unlikely to have resulted from photoadaptation because *Synechococcus* cells had photosynthetic characteristics typical of high light populations¹⁰. This observation supports the argument that phycoerythrin may have a nitrogen storage function in this organism⁸.

Integrated primary productivity on 21 July was three times higher than previous summer measurements in the same area¹², and was primarily the result of high photosynthetic rates in the *Synechococcus* population. The maximum rate of 142 fg C cell⁻¹ d⁻¹ on 21 July, corresponds to a specific growth

Table 1 Pigment concentrations, primary productivity and cell numbers integrated over the upper 25 m between 20 and 25 July 1986 in the northern Sargasso Sea, at 32°41' N, 63°01' W

		July 1986			
		20	21	22	25
Pigments ($\mu\text{g m}^{-2}$)					
Phycocerythrin					
<i>Synechococcus</i> spp.	(>0.6 μm)	9,110	—	7,340	1,850
Chlorophyll <i>a</i>					
<i>Synechococcus</i> spp.	(0.2–0.6 μm)*	23	26	6	—
	(0.6–1 μm)	490	789	465	285
	(1–5 μm)*	146	406	77	—
Total <i>Synechococcus</i> spp.	(>0.2 μm)	659	1,221	548	—
Total algae	(>0.2 μm)*	1,071	869	892	—
All phytoplankton	(>0.2 μm)	1,730	2,090	1,440	880
Primary productivity ($\text{mg C m}^{-2} \text{d}^{-1}$)					
<i>Synechococcus</i> spp.	(0.2–0.6 μm)*	—	1	1	—
	(0.6–1 μm)	—	25	16	—
	(1–5 μm)*	—	14	3	—
Total <i>Synechococcus</i> spp.	(>0.2 μm)	—	40	20	—
Total algae	(>0.2 μm)*	—	45	46	—
All phytoplankton	(>0.2 μm)	—	85	66	—
Cell numbers ($\text{cells m}^{-2} \times 10^{-9}$)					
<i>Synechococcus</i> spp.	(0.2–0.6 μm)	10	7	5	—
	(0.6–1 μm)	215	210	358	} 208
	(1–5 μm)	64	108	59	
Total <i>Synechococcus</i> spp.	(>0.2 μm)	289	325	422	

* Calculated.

rate (μ) of 1.2 d^{-1} and a 14 h generation time, assuming a mean cell size of $1 \times 0.8 \mu\text{m}$ [$0.5 \mu\text{m}^3$] and $121 \text{ fg C } \mu\text{m}^{-3}$ (ref. 13). (Waterbury *et al.*¹⁴ reported $210 \text{ fg C cell}^{-1}$ for an optimally growing Sargasso-Sea *Synechococcus* isolate, WH7803, which can reach dimensions of $2.2 \times 1.0 \mu\text{m}$ (1.0 μm diameter) which corresponds to $122 \text{ fg C } \mu\text{m}^{-3}$.) This estimated growth rate is within the range of maximum rates for laboratory cultures of marine *Synechococcus* (0.5 – 1.7 d^{-1}) (refs 14 and 15). *Synechococcus* photosynthesis dropped from 142 to $69 \text{ fg C cell}^{-1} \text{ d}^{-1}$ by 22 July, and the decline coincided with the main increase in cellular abundance.

Synechococcus species increased from 289 – $422 \times 10^9 \text{ cells m}^{-2}$ between 20 and 22 July, and there were 2–5 times more *Synechococcus* cells in the upper 25 m than previously observed in the Sargasso Sea in summer¹². Also, *Synechococcus* size distribution was shifted towards larger cells on 21 July, which is more typical of those from oceanic and neritic nitraclines^{9,12} and suggests that the surface population was nitrogen sufficient on the 21 July.

The difference in nitrate concentrations in the upper water

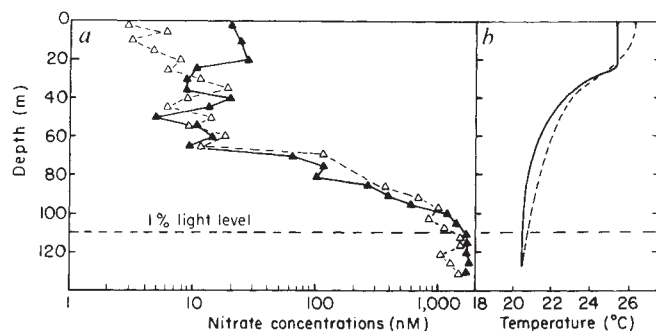


Fig. 2 Vertical profiles of nitrate concentrations and temperature were obtained from a submersible pump profiling system in the northern Sargasso Sea at 32°42' N, 63°01' W. The chemiluminescent technique⁷ was used for samples from 0–65 m and the colorimetric method⁶ for samples below 65 m. Graph *a* shows nitrate distributions on 20 July ($\blacktriangle$ — $\blacktriangle$) and 25 July ($\triangle$ — $\triangle$) and graph *b* shows temperature profiles on 20 July (—) and 25 July (---).

column between 20 and 25 July was $417 \mu\text{mol m}^{-2}$. This nitrate could support a total new production of 30 mg C m^{-2} , or 492×10^9 *Synechococcus* cells m^{-2} , assuming $61 \text{ fg C cell}^{-1}$ (refs 13 and 14) and a C:N ratio of 6:1 (ref. 14). In fact, the increase in *Synechococcus* numbers between 20 and 22 July was only 133×10^9 cells m^{-2} , or 27% of the potential yield. This resulted from close coupling between production and grazing¹⁶ during the transient bloom, as numbers of flagellates doubled from 1 – 2×10^9 cells m^{-2} between 20 and 23 July. Grazers removed 70% of *Synechococcus* cellular production on 23 July.

The stoichiometry of nitrate disappearance, net cell production and grazing, are internally consistent with observations being made on a single water column and heat budget calculations are consistent with the warming that occurred in the upper 25 m (Fig. 2*b*). Consequently, our data support the concept of a nitrate-dependent transient bloom. Compared to previous observations in surface Sargasso Sea water, higher nitrate concentrations, higher *Synechococcus* standing stock, a cell-size distribution shifted towards larger cells, and higher cellular rates of photosynthesis were all observed during this event.

Our data provide little insight into the sources or supply mechanisms that resulted in the high nitrate concentrations on the 20 July. If the nitrate was supplied from the nitracline, diffusional mechanisms^{2,11} are precluded by the vertical nitrate gradient, but larger scale processes may have been involved¹⁷ and our observations were preceded by major storms. Surface nitrate concentrations may be increased by rainfall¹¹, but the rain preceding the monitoring period (Fig. 1*a*) probably could not have contributed enough nitrate¹⁸.

Transient nanomolar increases in surface layer nitrate concentrations do occur in stratified oceanic water columns and can selectively and rapidly induce *Synechococcus* growth to near maximum rates. Stoichiometry suggests that they alone responded to the elevated nitrate concentrations and this new production was rapidly transferred to grazers. These observations are consistent with the view that oceanic phytoplankton can exhibit high growth rates^{3,13,15}, and that rates of photosynthesis and growth of oceanic phytoplankton can change rapidly over short timescales (hours to days)⁵.

This paper is dedicated to Alison Rollins, assistant engineer, RV *Endeavor*, who died on 11 August 1986. We thank W.

Connelly, A. Langeley, N. A. Matlick, N. Nelson, O. Schofield, A. E. Smith, D. Steinberg, A. C. Steiner, B. Ver Hoven, G. Campbell and the RV *Endeavor* crew for assistance with measurements and M. R. Lewis, J. J. Cullen and C. S. Yentsch for reviewing the manuscript. This work was supported by the NSF and the NERC (UK).

Received 8 June; accepted 19 November 1987.

1. Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677-680 (1979).
2. Lewis, M. R., Harrison, W. G., Oakey, N. S., Herbert, D. & Platt, T. *Science* **234**, 870-873 (1986).
3. Goldman, J. C. in *Flows of Energy and Materials in Marine Ecosystems: Theory and Practice* (ed. Fasham, M. J. R.) 137-170 (Plenum, New York, 1984).
4. Jenkins, W. J. & Goldman, J. C. *J. Mar. Res.* **43**, 465-491 (1985).
5. Platt, T. & Harrison, W. G. *Nature* **318**, 55-58 (1985).
6. Strickland, J. D. H. & Parsons, T. R. *Fish Res. Bd Can. Bull.* **167**, 310.
7. Garside, C. *Mar. Chem.* **11**, 159-167 (1982).
8. Wyman, M. & Carr, N. G. *Limnol. Oceanogr.* (in the press).
9. Glover, H. E., Campbell, L. & Prézélin, B. B. *Mar. Biol.* **91**, 193-203 (1986).
10. Prézélin, B. B., Putt, M. & Glover, H. E. *Mar. Biol.* **91**, 205-217 (1986).
11. Garside, C. *Deep Sea Res.* **32**, 723-732 (1985).
12. Glover, H. E., Smith, A. E. & Shapiro, L. *J. Plank. Res.* **7**, 519-535 (1985).
13. Iturriaga, R. & Mitchell, B. G. *Mar. Ecol. Prog. Ser.* **28**, 291-297 (1986).
14. Waterbury, J. B., Watson, S. W., Valois, F. W. & Franks, D. G. in *Photosynthetic Picoplankton* (eds Platt, T. & Li, W. K. W.) *Can. Bull. Fish. aquat. Sci.* **214**, 71-120 (1986).
15. Campbell, L. & Carpenter, E. J. *Mar. Ecol. Prog. Ser.* **32**, 139-148 (1986).
16. Campbell, L. & Carpenter, E. J. *Mar. Ecol. Prog. Ser.* **33**, 121-129 (1986).
17. Klein, P. & Coste, B. *Deep Sea Res.* **31**, 21-37 (1984).
18. Knap, A., Jickells, T., Pszeny, A. & Galloway, J. *Nature* **320**, 158-160 (1986).

Perception of shape from shading

V. S. Ramachandran

University of California, San Diego, La Jolla, California 92093, USA

The human visual system can rapidly and accurately derive the three-dimensional orientation of surfaces by using variations in image intensity alone¹⁻⁵. This ability to perceive shape from shading is one of the most important yet poorly understood aspects of human vision. Here we present several findings which may help reveal computational mechanisms underlying this ability. First, we find that perception of shape from shading is a global operation which assumes that there is only one light source illuminating the entire visual image. This implies that if two identical objects are viewed simultaneously and illuminated from different angles, then we would be able to perceive three-dimensional shape accurately in only one of them at a time. Second, three-dimensional shapes that are defined exclusively by shading can provide tokens for the perception of apparent motion, suggesting that the motion mechanism is remarkably versatile in the kinds of inputs it can use. Lastly, the occluding edges which delineate an object from its background can also powerfully influence the perception of three-dimensional shape from shading.

Our first experiment shows that the derivation of shape from shading incorporates the constraint that there is only one light source illuminating all (or most) of the image. Notice that the strong sense of depth observed in Fig. 1a is derived exclusively from shading. The sign of perceived depth, however, is ambiguous for the brain has no way of knowing what the direction of illumination is. Consequently, the objects in the display can be perceived either as being convex or concave. In preliminary experiments we found that when we perceptually 'reversed' the sign of depth for any one object in this display then all of the other objects reversed their sign as well. This is consistent with previous reports that when certain ambiguous visual figures are viewed simultaneously the brain seems to apply a global rather than a local strategy to resolve the ambiguity⁶.

These findings raise an interesting question. Is the propensity to see all the objects in Fig. 1a as simultaneously convex (or simultaneously concave) based on a tendency to assign identical depth values to all the objects or is it based on a tacit assumption

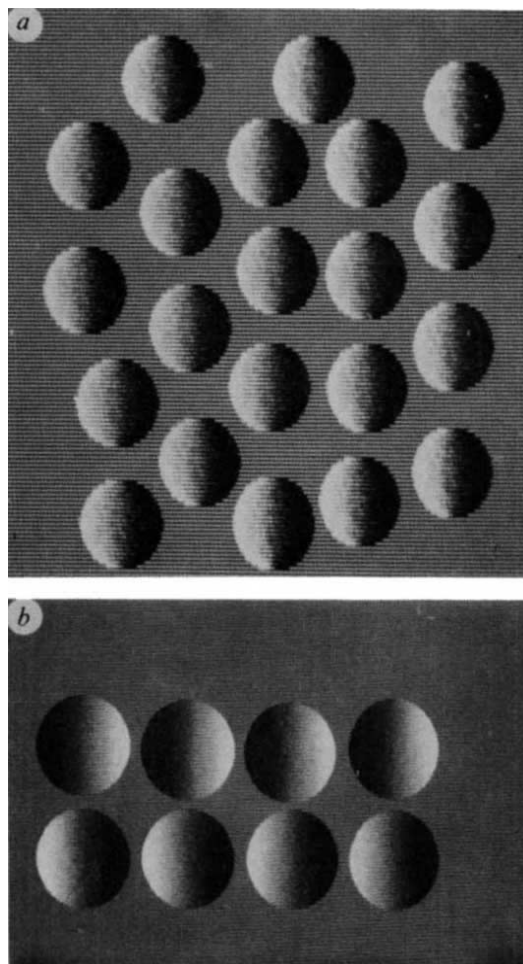
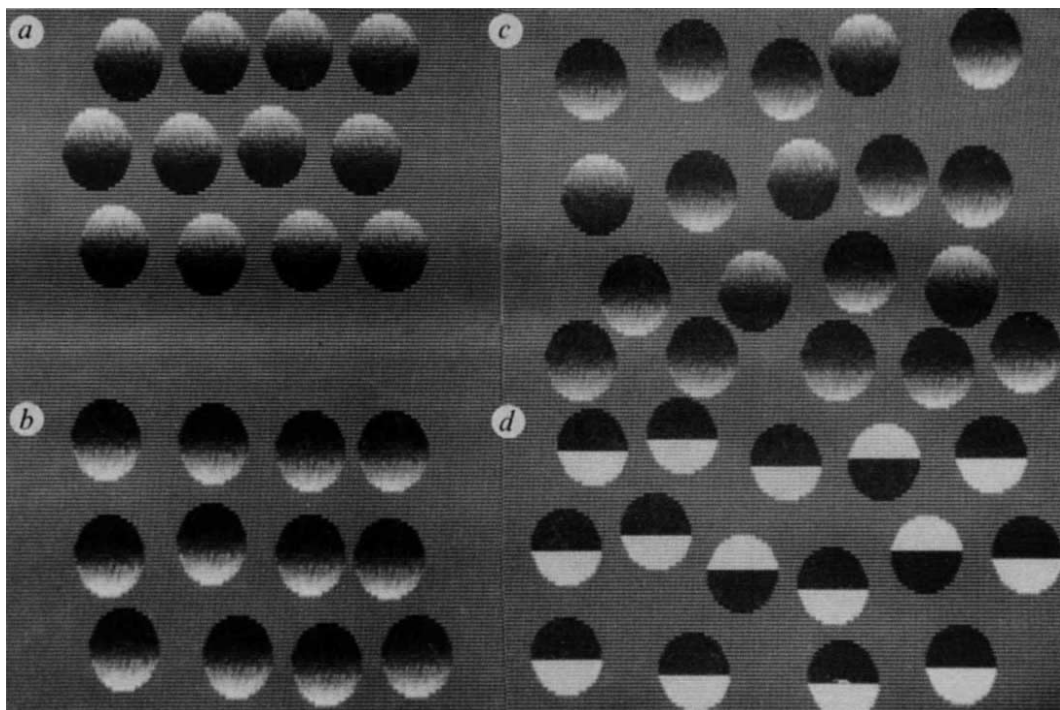


Fig. 1 a, Ambiguous objects that can be seen as either spheres or holes. Depth perception in this picture can be improved by squinting one's eyes to blur the image slightly. b, Demonstration of the 'single-light source' constraint. When one horizontal row is seen as convex the other row is always seen as concave. It is very difficult to see all objects as being simultaneously convex or simultaneously concave. This simple display shows that the brain will accept only one light source at a time for the entire visual image (or large portion of it).

that there is only a single light source illuminating the image? To find out we used a display (Fig. 1b) consisting of two rows of objects which were mirror images of each other. Eleven subjects viewed this display and reported that when one row of objects was convex then the mirror images were always concave (and vice versa). It was in fact almost impossible to see all objects as being simultaneously convex or concave. Notice that either of the two horizontal rows can be seen as concave or convex when the other row is covered with a piece of cardboard. When viewed simultaneously, however, seeing one row as convex forces the visual system to see the other row as concave. This is an important observation for it suggests that the brain prefers a 'common-light-source' assumption to a 'common-depth' assumption—a principle that actually violates the Gestalt law of common fate. This may be because our brains evolved in a solar system which has only one sun. Furthermore, our observations also contradict a common misconception shared by many psychologists⁷—that the brain always prefers the simplest interpretation of any visual pattern. In Fig. 1b it would be simpler to see all objects as either simultaneously convex or simultaneously concave but the visual system actually rejects these interpretations.

In Fig. 2 there is a strong tendency to perceive the objects in the top left panel (a) as convex and those in the bottom left

Fig. 2 *a*, These are usually seen as convex as the visual system automatically assumes that the source of illumination is at the top. Interestingly it is the orientation on the retina that matters—not the phenomenal or 'gravitation' vertical. The reader can verify this by viewing panel *a* while hanging upside down from a ceiling. *b*, These objects are usually seen as hollow but they can sometimes be perceptually reversed to generate convex objects. *c*, A mixture of objects selected from panels *a* and *b*. In this display it is very difficult perceptually to reverse any of the objects. A reversal can be obtained by turning the page upside down. Notice that all the convex shapes can be grouped together mentally so that they are segregated clearly from the concave shapes. This segregation is



easier to achieve for top-bottom differences in illumination than for left-right differences—a point that can be verified by rotating the page by 90°. *d*, 'Control' for *c*. The tokens used are similar to those in *c* in terms of luminance polarity (although their positions have been randomized) but they do not convey any depth. It is very difficult to segregate the tokens on the basis of differences in polarity, suggesting that the effects observed in *c* must have been due to three-dimensional shapes.

panel (*b*) as concave. This suggests that in addition to the unitary-light-source constraint there is also a tendency to assume that the light source is at the top; an effect that is well known to artists. The effect is especially strong if a mixture of such objects is presented (Fig. 2*c*). In this display it is also possible to group all the convex shapes together mentally to form a cluster that is clearly segregated from the background of concave shapes. This is a surprising result as it is usually assumed that only certain elementary stimulus features such as orientation, colour and 'terminators' can be grouped together in this way⁸⁻¹⁰. Figure 2*c* shows that even three-dimensional shapes that are conveyed by shading can provide tokens for perceptual grouping and segregation. To make sure that the effect was not due to some more elementary image feature (such as luminance polarity) we produced a control stimulus (Fig. 2*d*) in which the targets were similar to those in Fig. 2*c* in terms of luminance polarity but did not convey any depth. In this display, it is difficult to segregate the tokens on the basis of differences in polarity suggesting that the effects observed in Fig. 2*c* must have been based on three-dimensional shapes. Eleven naive subjects confirmed this striking difference between the two displays when they were asked to compare them directly. Segregation is also much more more pronounced for top-down differences in illumination than for left-right differences. For instance, if Fig. 2*c* is rotated by 90° perceived depth becomes less compelling and the degree of segregation is also reduced correspondingly. This further supports the view that effect depends on the three-dimensional shapes of the tokens rather than on luminance polarity.

Our next experiment shows that shapes that are defined exclusively by shading can provide an input to apparent motion. To show this we began with the two objects, a convex one, and a concave one, presented simultaneously one below the other in the first frame (Fig. 3*a*). This was switched off and replaced with the second frame in which the two objects were simply made to exchange locations (Fig. 3*b*). The frames are shown

side by side for clarity but in the original experiment they were optically superimposed and alternated rapidly. The objects were 1.5° wide and they were separated by a distance of about 2°. A very vivid impression of apparent motion could be obtained from this stimulus sequence. Eleven naive subjects reported that they could see a sphere jumping up and down as it alternately filled and vacated two holes in the background. We varied the stimulus onset asynchrony (SOA) over a wide range, keeping the inter-stimulus interval constant at zero ms and found that apparent motion disappeared when the SOA was less than ~75 msec. Similar effects were observed when we used cylindrical targets (Fig. 3*c* and *d*) instead of circular ones. Interestingly, with cylindrical targets the depth often took 30–60 s to

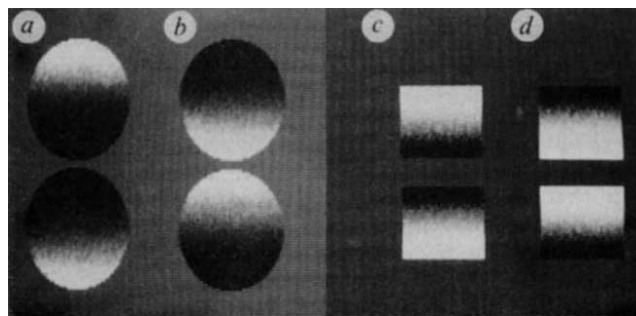


Fig. 3 *a* and *b*, Two frames of an apparent motion sequence. In the first frame (*a*) there is a sphere on top and a 'hole' just below it. In the second frame (*b*) the two figures have exchanged places. The two frames are shown side-by-side for clarity but in the original experiment they were optically superimposed and alternated. This created the striking impression of a sphere moving vertically as it 'filled' and vacated two sockets in the background. *c* and *d* Show cylinders rather than spheres (see text). Several seconds may elapse before depth emerges in these displays.

emerge and during this long latency no apparent motion could be perceived. Once depth became clearly defined, however, the motion of the cylinder became instantly visible. We conclude that the perception of apparent motion must necessarily be preceded by a computation of three-dimensional shape from shading based on the single-light-source constraint. Certain cells in the middle temporal area of monkey cortex¹¹ are known to respond to apparent motion of texture borders¹² and it might be interesting to see if these cells would also respond to motion that is derived from shape perceived from shading.

Is the apparent motion seen in Fig. 3 based on shape from shading *per se* or some other more elementary image feature? It is conceivable that similarity of brightness polarity could have served as a basis for establishing correspondence in these displays. This seems unlikely as the depth impression often took 30–60 s to develop and apparent motion was never perceived until depth emerged. To show the point more directly we also used a 'control' display which conveyed no depth but was similar to Fig. 3 in terms of brightness polarity (the targets resembled the controls shown in Fig. 2d). When the two frames were alternated, random, incoherent flicker was seen and, unlike Fig. 3, there was never any impression of unambiguous vertical apparent motion. This implies that motion correspondence in Fig. 3 was being established between three-dimensional shapes rather than more primitive image features.

In deriving shape from shading the visual system can also take advantage of information provided by the occluding edges that delineate an object from its background¹. We constructed Fig. 4a and b to show this principle. The luminance gradients along the horizontal axis are identical for both images (a and b) yet the surfaces look very different. Figure 4a looks like three metal cylinders touching each other, whereas Fig. 4b conveys unmistakably the impression of a corrugated metal sheet. Notice that the shape of each surface faithfully follows the contours at the upper and lower borders of the object. These contours seem to propagate inward to influence the perceived depth of the entire surfaces. Interestingly, there is also a corresponding shift in the perceived direction of illumination. In Fig. 4a the light source occupies a position exactly normal to the surface, that is, it has the same location as the observer. In Fig. 4b, on the other hand, the light seems to come from the left. It is noteworthy that such striking changes in perception can be produced simply by altering the outlines of the objects.

Even an illusory outline will produce this effect in Fig. 4c which we generated by superimposing an illusory circle^{13,14} on a simple one-dimensional luminance gradient. This figure initially looks quite flat but on prolonged viewing the region corresponding to the illusory circle starts bulging out towards the observer and may even detach itself from the background to take on the appearance of a sphere. Again, this observation shows that the derivation of shape from shading can be strongly influenced by segmentation boundaries. The important role played by segmentation boundaries has also been shown by us previously for a variety of other visual capacities such as stereopsis¹⁵, apparent motion⁶ and three-dimensional structure from motion¹⁶.

Our last experiment shows the role of top-down influences in the perception of three dimensional shape from shading. We found that when the objects were arranged in regular rows (Fig. 4d) it was sometimes possible to 'imagine' the display as a set of 16 holes in an opaque grey occluder through which we could see two cycles of a sine-wave grating¹⁷. The perceptual switch was quite compelling, and when it occurred the impression of curved three-dimensional shapes vanished completely. This shows that there is a direct and powerful interaction between occlusion cues and the derivation of shape from shading and also implies that the process can be strongly modulated by top-down influences such as visual imagery.

We find that the visual system's ability to determine the three dimensional shape of a shaded region is strongly constrained

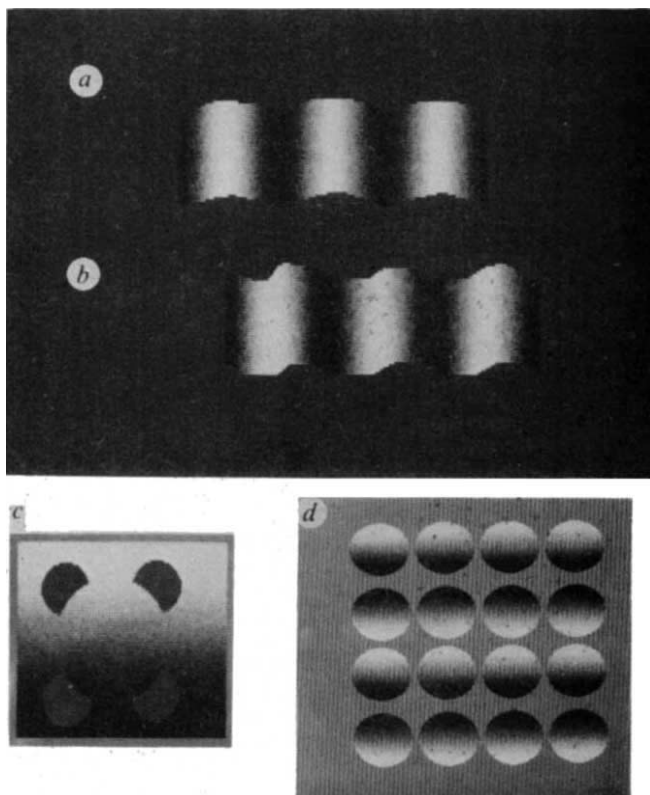


Fig. 4 a and b, Here there are identical luminance gradients along the horizontal axis yet the surfaces look very different. In fact the surfaces seem faithfully to follow the shapes of the occluding edges at the top and bottom of each figure. The depth effect can take as long as a minute to evolve. It would help to squint your eyes slightly and to view the figures from a distance. c, An 'illusory' circle superimposed on a simple one-dimensional luminance gradient. The area enclosed by the circular outline bulges out to assume the appearance of a spherical object. We find this effect (as well as the 'spheres' observed in Fig. 2) can be obtained even if the luminance gradient is not a cosine function of the kind that would be associated with real spherical surfaces. d, This display illustrates that top-down influences can play a role in the perception of shape from shading. The display can be seen either as alternate horizontal rows of convex and concave objects or as two cycles of a sine-wave grating viewed through 16 'holes' cut out of a grey opaque occluder. The effect is easier to obtain if the page is rotated through 90°.

by the assumption that there is only one light source in the image and also by the shape of the occluding edges that delineate this area from the surround. Furthermore, once these three-dimensional shapes have been made explicit, the system can use them for a variety of visual skills such as figure-ground segregation and motion perception.

We thank Y. Fisher, D. Kleffner, D. Rogers-Ramachandran, S. Cobb, J. A. Deutsch, H. Pashler and F. H. C. Crick for their helpful comments. V. S. Ramachandran was supported by biomedical and academic senate research funds from the University of California.

Note added in proof: We find that only luminance-derived outlines can support the perception of three-dimensional shapes. If the object's outline is conveyed exclusively by chromatic borders, the impression of solidity disappears completely.

Received 6 August; accepted 2 November 1987.

1. Witkin, A. P. & Tenenbaum, J. M. in *Human Machine Vision* (eds Beck, J., Hope, B. & Rosenfeld, A.) 481–543 (Academic, New York, 1983).
2. Pentland, A. J. *opt. Soc. Am.* **72**, 448–455 (1982).
3. Todd, J. & Mingolla, E. *J. exp. Psychol.* **9**, 583–595 (1983).
4. Horn, B. K. P. in *The Psychology of Computer Vision* (ed. McGraw Hill, New York, 1975).
5. Yonas, A., Kuskowski, M. & Sternfels, S. *Child Dev.* **50**, 495–500 (1979).

6. Ramachandran, V. S. & Anstis, S. M. *Scient. Am.* **254**, 102-109 (1986).
7. Restle, F. in *Organization and representation in perception* (ed. Beck, J.) 31-36 (Lawrence Erlbaum, London, 1982).
8. Treisman, A. *Scient. Am.* **255**, 114-126 (1986).
9. Julesz, B. *Foundations of Cyclopean Perception* (Chicago University Press, 1971).
10. Beck, J. *Percept. Psychophys.* **1**, 300-302 (1966).
11. Albright, T. *Soc. Neurosci. Abstr.* **13**, 1626 (1987).
12. Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. *Vision Res.* **13**, 1399-1401 (1973).
13. Gregory, R. L. *Nature* **238**, 51-52 (1972).
14. Kanizsa, G. *Scient. Am.* **234**, 48-52 (1976).
15. Ramachandran, V. S. *Percept. Psychophys.* **39**, 361-373 (1986).
16. Ramachandran, V. S., Cobb, S. & Rogers-Ramachandran, D. *Soc. Neurosci. Abstr.* **13**, 630 (1987).
17. Campbell, F. W. & Robson, J. G. *J. Physiol., Lond.* **197**, 551-556 (1968).

Real-time imaging of evoked activity in local circuits of the salamander olfactory bulb

John S. Kauer

Departments of Neurosurgery, and Anatomy and Cell Biology,
New England Medical Center and Tufts Medical School,
Boston, Massachusetts 02111, USA

The encoding of olfactory information in the central nervous system (CNS) depends on spatially distributed patterns of activity generated simultaneously in many neuronal circuits¹⁻⁴. Optical neurophysiological recording permits analysis of neural activity non-invasively and with high spatial and temporal resolution⁵. Here, a video method for imaging voltage-sensitive dye fluorescence *in vivo* is used to map neuronal activity in local circuits of the salamander olfactory bulb. The method permits the imaging of simultaneous ensemble transmembrane activity in real time. After electrical stimulation of the olfactory nerve, activity spreads centripetally from the sites of synaptic input to generate non-homogeneous response patterns that are presumably mediated by local circuits within the bulbar layers. The results also show the overlapping temporal sequences of activation of cell groups in each layer. The method thus provides high resolution, sequential video images of the spatial and temporal progression of transmembrane events in neuronal circuits after afferent stimulation and offers the opportunity for studying ensemble events in other brain regions.

The design of the recording equipment and the planar laminae of the salamander olfactory bulb are shown in Fig. 1a and b. These layers are oriented approximately perpendicular to the dorsal bulbar surface, in contrast to the concentric organization in the mammal, and can be identified by sequential recording of electrically-evoked field potentials⁶, characteristic of each layer, across the dorsal surface.

A highly simplified scheme of the synaptic organization of the bulb is shown in Figs 1b and 3a. Mitral/tufted and periglomerular cells receive excitatory input from afferent receptor axons. There are reciprocal dendrodendritic synapses between excitatory mitral/tufted and inhibitory periglomerular cells, and between mitral/tufted and inhibitory granule cells⁷. The connectivity among these widespread negative-feedback circuits, is thought to mediate the long-lasting periods of inhibition seen with electrical and odour stimulation. Electrophysiological^{4,8-12} and 2-deoxyglucose¹³ analyses have shown that stimulation with an odorant gives rise to spatially and temporally non-uniform patterns of activity, presumably shaped by these local bulbar circuits^{9,14}.

To characterize these ensemble response patterns, changes in the fluorescence of the voltage-sensitive dye RH414 (ref. 15) were observed at video rates. This allowed the study of bulbar responses elicited by electrical shocks, taking advantage of the more synchronous activation these stimuli provide over odorant pulses¹⁶. The tiger salamander (*Ambystoma tigrinum*) is a useful animal⁸ for these studies as electrophysiological recordings have

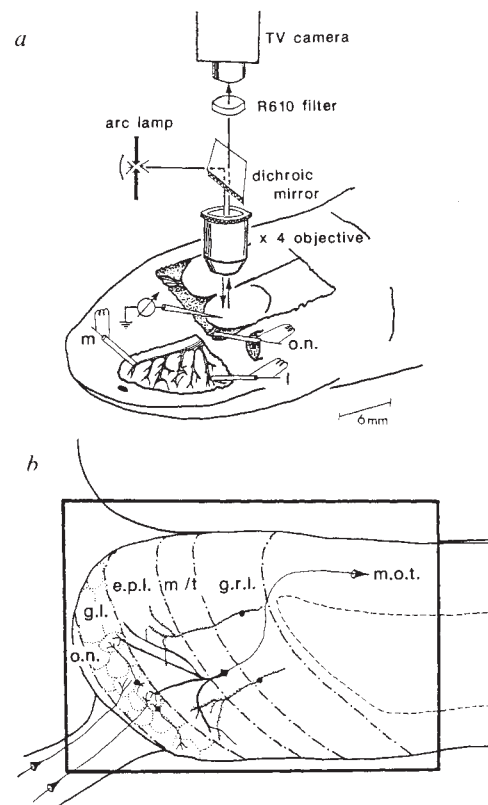
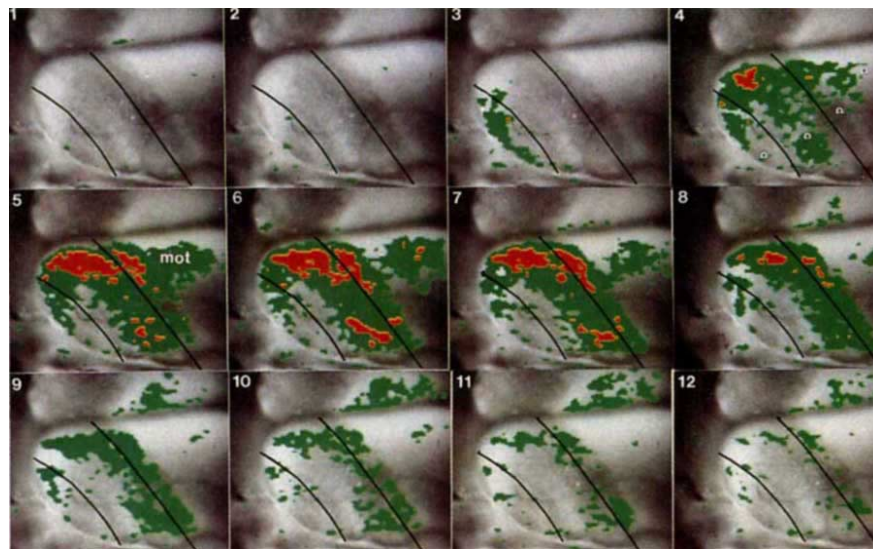


Fig. 1 The design of the dye-recording equipment showing light source, dichroic mirror, microscope objective, filter, video camera and recording and stimulating electrodes (m, medial; l, lateral; o.n., olfactory nerve). **b**, Dorsal view of the olfactory bulb laminae schematically showing the organization of the major cell types (g.l., glomerular; e.p.l., external plexiform; m/t, mitral/tufted cell body; g.r.l., granule-cell body layers; m.o.t., medial olfactory tract). **Methods.** The animal and the electrodes were held on the stage of an Olympus BHS-2 microscope. The brain was epi-illuminated by a Kratos Analytical Instruments (Ramsey, New Jersey) 150 W xenon-mercury arc lamp with a feedback-stabilized power supply. A dichroic mirror reflected wavelengths below 580 nm, exciting RH414 ($\lambda_{\text{max}} = 530$ nm), and transmitted fluorescence at 610 nm through an Olympus R610 nm filter to a Dage-MTI (Michigan City, Illinois), series-68 television camera with a Newvicon tube. Sequences of 16 video frames taken at 30 frames s^{-1} (128 pixels $\times$ 128 pixels $\times$ 8 bits) were digitized and processed by an Imaging Technology (Woburn, Massachusetts) FG100-AT board in a Compaq 286 Deskpro computer (Houston, Texas). Thirty-eight land-phase animals were anaesthetized with ketamine HCl (10 mg per 100 g administered every 45-60 min), immobilized with *d*-tubocurarine (0.7 mg per 100 g), and locally anaesthetized with lidocaine at the wound and head fixation points. The voltage-sensitive dye RH414 (ref. 15) (1.5 mg ml^{-1} amphibian Ringer) was placed on the exposed bulb for 20-25 min; excess dye was washed off with Ringer solution. No other voltage-sensitive dyes were tested. There was no obvious deterioration in the fluorescence or electrical signals during the 4-6 h experiments. Electrical stimulation (700 μs duration, 2-5 $\times$ threshold) was carried out by a concentric bipolar electrode placed on the trunk of the olfactory nerve at its entry to the olfactory bulb or on dorso-lateral or dorso-medial fascicles overlying the nasal sac (see Fig. 1a). Field potentials were recorded on the surface of the bulb (see Fig. 3b). The video field shown in the frames of Fig. 2 is delineated by heavy black lines in Fig. 1b.

been made from single receptor¹⁷ and bulbar^{8,9} cells, and voltage-sensitive dye signals have been recorded with a diode array^{18,19}.

After preparing the animal as in Fig. 1, changes in RH414 fluorescence were monitored by acquiring pairs of 16-frame (33 ms frame⁻¹) sequences. Each experimental run consisted of subtracting a 16-frame sequence without the stimulus, from a

Fig. 2 Sequence of 12 video-rate frames showing the fluorescence changes superimposed on bright-field image of the bulb after a shock to the olfactory nerve delivered between frames 1 and 2; dots in frame 4 indicate sites in all frames from which the plots in Fig. 3b were made. Green, activity one standard deviation (s.d.) above the pre-stimulus baseline; red, activity two s.d.s above baseline. Data were 'smoothed' with a low-pass spatial digital filter. Black line over the bulb to the left side of each frame indicates the border between glomerular and external plexiform layers. The line to the right is roughly over the mitral-body layer. This run is an average of 10 sequence-pair subtractions. Medial olfactory tract, m.o.t.



sequence with the stimulus, pixel by pixel, to generate net changes in fluorescence in each frame. Using this paradigm, stimulus-evoked signals were seen in all animals, often after only a single trial, although the noise was reduced by averaging 3–10 runs. The signals seen in each trial were large²⁰, ranging from 0.5 to 1% of the background fluorescence, consistent with signals recorded with a diode array^{18,19}.

The sums of all background/stimulus-pair subtractions were normalized to compare data across experiments. In Fig. 2 the colours represent signals one (green) and two (red) standard deviations above the pre-stimulus baseline. Final data were superimposed on the bright-field image of the bulb to correlate the responses directly with the cell laminae. The data are displayed here as series of photographs, but their spatial and temporal characteristics are best seen by displaying them as a short 'movie'.

No signals were observed under the following conditions: averaging many trials without the dye; subtracting sets of sequences from one another, both without stimulus shocks; observing activity before the shock in experimental trials (Fig. 2, frame 1); and using a non-voltage-sensitive dye (1 mg ml⁻¹ sodium fluorescein) in place of RH414. Both electrical and fluorescence responses were abolished by 10 μ M tetrodotoxin (TTX) placed on the bulb. The signal latency increased when the propagation delay was lengthened by moving the stimulating electrode to a peripheral olfactory nerve fascicle. No changes in fluorescence, except bleaching, were observed when RH414 was observed using a 500 nm cutoff dichroic mirror. These results provide evidence that the signals arose from dye-related neurophysiological events.

The 12-frame sequence after a shock to the olfactory nerve trunk in Fig. 2 shows in detail the spread of depolarizing potentials within the bulbar layers: initially into the glomerular layer (frame 3)—the site of olfactory nerve termination; next into the external plexiform layer (frame 4)—the site of dendro-dendritic interactions between mitral/tufted secondary dendrites and granule-cell spines; for a brief time (frames 5–7), out along the medial olfactory tract—the position of output mitral/tufted axons leading to more central olfactory relays; and then remaining within the external plexiform layer, mitral body, and granule cell layers (frames 8–12) for some 160 ms after activation of the medial olfactory tract. There is no response in the olfactory nerve because the short 1–2 ms latency of the nerve-potential, propagating 100 μ m at 0.1 m s⁻¹, is too brief for the present method to resolve. Signals could be seen propagating into the contralateral bulb (frames 7–12), characteristically 4 frame times after the ipsilateral signal.

As much is known about the synaptic organization of the

olfactory bulb, the present findings can be correlated with the connections among the cells and with other physiological data. Figure 3a shows the major synaptic connections among these elements. Figure 3b shows a comparison of the signal time courses obtained from each lamina using fluorescence and electrophysiological methods. The rising phases of signals obtained with the present method (solid lines) have time courses similar to those from analogous regions measured with the diode array (dashes). As the video method detects sustained changes in dye fluorescence, it appears to register the long falling phase of the signals more faithfully than the capacity-coupled diode detectors which, for the data shown here, had a time constant of 0.1 s (ref. 18).

The peaks of the fluorescent signals in the glomerular layer in Fig. 3b correlate in time with depolarization (*) of mitral/tufted cells seen in the field-potential and intracellular traces, and with depolarization of the periglomerular cells. The longer latency peaks of the fluorescent signal in the external plexiform layer and the granular cell layer correlate with depolarization of the granule dendrites and cell bodies after their activation via the mitral/tufted cells. The longer falling phase of the video signal may reflect the known long-term activity in the granule cells^{7,9,14}. The fluorescent signal in the medial olfactory tract has a shorter latency than the external-plexiform-layer and granular-cell-layer signals. The time course of this axonal activity is analogous to the brief burst and curtailment of mitral/tufted output spikes seen in single-unit records after orthodromic stimulation^{6–9,14}, which is thought to be generated by inhibition from external plexiform-layer circuits (see intracellular trace in Fig. 3b m/t.b.l.). Although there is profound hyperpolarization of mitral/tufted cells following the spike after orthodromic stimulation as shown in the intracellular trace, this is presumably not seen in the fluorescence record because it is swamped by the depolarization of the much more numerous granule cells.

Additional experiments have provided information on the distribution of bulbar responses obtained after electrical shocks to medial and lateral bundles of olfactory nerve fascicles (see Fig. 1a for electrode placements). In these experiments, stimulation of each fascicle elicited small field potentials, but large, widespread changes in dye fluorescence, resembling those shown in Fig. 2. Stimulation of either bundle elicited activity across much of the bulb, even though the stimuli were delivered to nerves coming from different regions of the nasal cavity. These data imply that stimulation of local epithelial sites can influence widespread regions of the bulb and provide physiological evidence for activity in the divergent connections between restricted epithelial loci and the bulb mapped with horseradish

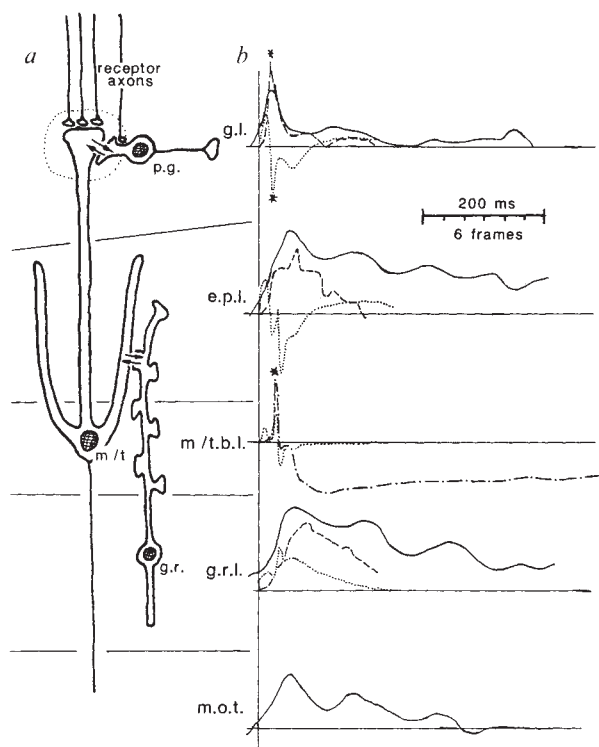


Fig. 3 *a*, Highly schematic diagram showing connections between major cell types. p.g., Periglomerular cells; e.p.l., external plexiform layer; m/t.b.l., mitral/tufted body layer; g.r.l., granule layer; m.o.t., medial olfactory tract. *b*, Time-course comparison of signals from different bulbar laminae. The positions of the pixels across all frames, whose amplitude values were used for the video plot are shown in Fig. 2, frame 4. Amplitudes have been scaled to about the same value for this comparison. Diode-array signals are from Orbach and Cohen¹⁸. Fluorescent signals obtained with a diode array (---); fluorescent signals observed by the video method (—); field-potential signals recorded with a surface-recording pipette (.....); and an intracellular record from a mitral/tufted cell body (-.-.-).

peroxidase in this animal²¹ and in the rat²². The network built up by such overlapping divergent connections is presumably important for generating odour-elicited activity patterns^{1,2,21,22}.

Are the fluorescence signals from neurons, glia, or other tissues such as blood vessels? No signals were correlated with blood vessels, which were easily seen in the bright-field images generated for each dye record. A previous study¹⁸ using a diode array to measure RH414 after olfactory-nerve stimulation found a rapid fluorescent signal which correlated with the compound action potential from the olfactory nerve layer. This suggested that RH414 registered action-potential activity in the olfactory nerve, not from its surrounding glial sheath. Although this first peak was too fast to be detected with the present system, the video signals from the other laminae correlated well with the diode-array records and the electrophysiological recordings, with the shortest detectable signal lasting about as long as one frame as in Fig. 3b, glomerular-layer record. All signals recorded in this study were faster than the dye-independent, intrinsic signals observed by Grinvald *et al.*²³ in rat somatosensory cortex and the signals recorded by Blasdel and Salama²⁴ using a video system applied to monkey visual cortex. The small size of the signals from monkey cortex required averaging 1,800 frames, precluding observation of the temporal characteristics of the response. The present signals are also faster than those seen in amphibian glia²⁵ and elasmobranch cerebellum²⁶. Taken together, these observations suggest that the signals in this study were recorded from neural structures.

These experiments have provided the first combined topographical and real-time *in vivo* neuronal recordings from a

vertebrate CNS structure by video-rate imaging of a voltage-sensitive dye. Detailed reconstruction of the dye signals in three dimensions, the use of pharmacological blockers of synaptic activity, and the generation of responses to odours should provide detailed information on the mechanisms which underlie the encoding of olfactory information. Although the present use of the technique provides high spatial resolution, it should be possible to improve temporal response by using faster video-frame acquisition rates. The method should be valuable for studying other distributed processes in neural circuits such as sensory-motor integration, learning, memory and epilepsy.

I thank Jeffrey Lu, Robert Winter and Dr Samuel Neff for assistance with computer programming, Dr K. Hamilton for the intracellular record, Drs H. Orbach and L. Cohen for the diode-array signals, and Dr A. Grinvald for suggesting the non-voltage sensitive dye control and for providing amphibian RH414. Supported by a US Public Health Service grant and by the Department of Neurosurgery, Tufts-New England Medical Center.

Received 22 September; accepted 23 November 1987.

1. Kauer, J. S. in *Neurobiology of Taste and Smell* (eds Finger, T. E. & Silver, W. L.) 205-231 (Wiley, New York, 1987).
2. Shepherd, G. M. in *Taste, Olfaction and the Central Nervous System* (ed. Pfaff, D. W.) 307-332 (Rockefeller, New York, 1985).
3. Gesteland, R. C., Lettvin, J. Y. & Pitts, W. H. *J. Physiol., Lond.* **181**, 525-559 (1965).
4. Doving, K. B. *Acta physiol. scand.* **130**, 285-298 (1987).
5. Cohen, L. B., Salzberg, B. M. & Grinvald, A. *Rev. Neurosci.* **1**, 171-182 (1978).
6. Rall, W. & Shepherd, G. M. *J. Neurophysiol.* **31**, 884-915 (1968).
7. Shepherd, G. M. *Physiol. Rev.* **52**, 864-917 (1972).
8. Kauer, J. S. *J. Physiol., Lond.* **243**, 695-715 (1974).
9. Hamilton, K. A. & Kauer, J. S. *Brain Res.* **338**, 181-185 (1985).
10. Harrison, S. A. & Scott, J. W. *J. Neurophysiol.* **56**, 1571-1589 (1986).
11. Macrides, F. & Chorover, S. *Science* **175**, 84-87 (1972).
12. Meredith, M. & Moulton, D. G. *J. gen. Physiol.* **71**, 615-643 (1978).
13. Stewart, W. B., Kauer, J. S. & Shepherd, G. M. *J. comp. Neurol.* **185**, 715-734 (1979).
14. Mori, K. *Prog. Neurobiol.* **29**, 320-375 (1987).
15. Grinvald, A., Anglister, L., Freeman, J. A., Hildesheim, R. & Manker, A. *Nature* **308**, 848-850 (1984).
16. Kauer, J. S. & Shepherd, G. M. *Brain Res.* **85**, 108-113 (1975).
17. Getchell, T. V. & Shepherd, G. M. *J. Physiol., Lond.* **282**, 521-540 (1978).
18. Orbach, H. S. & Cohen, L. B. *J. Neurosci.* **3**, 2251-2262 (1983).
19. Kauer, J. S., Senseman, D. & Cohen, L. B. *Brain Res.* **418**, 255-261 (1987).
20. Cohen, L. B. & Salzberg, B. M. *Rev. Physiol. Biochem. Pharmac.* **83**, 35-88 (1978).
21. Kauer, J. S. *Anat. Rec.* **200**, 331-336 (1981).
22. Stewart, W. B. & Pedersen, P. E. *Brain Res.* **411**, 248-258 (1987).
23. Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D. & Wiesel, T. N. *Nature* **324**, 361-364 (1986).
24. Blasdel, G. G. & Salama, G. *Nature* **321**, 579-585 (1986).
25. Konnerth, A. & Orkand, R. *Neurosci. Lett.* **66**, 49-54 (1986).
26. Konnerth, A., Obaid, A. L. & Salzberg, B. M. *Biol. Bull.* **169**, 553 (1985).

Haemodynamic shear stress activates a K⁺ current in vascular endothelial cells

Søren-Peter Olesen*, David E. Clapham†‡ & Peter F. Davies*

* Vascular Research Division, Department of Pathology, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

† Cardiovascular Division, Department of Medicine, Brigham & Women's Hospital, and Department of Physiology and Biophysics, Harvard Medical School, Boston, Massachusetts 02115, USA

The endothelial lining of blood vessels is subjected to a wide range of haemodynamically-generated shear-stress forces throughout the vascular system¹. *In vivo* and *in vitro*, endothelial cells change their morphology^{2,3} and biochemistry⁴ in response to shear stress in a force- and time-dependent way, or when a critical threshold is exceeded^{5,6}. The initial stimulus-response coupling mechanisms have not been identified, however. Recently, Lansman *et al.*⁷ described stretch-activated ion channels in endothelial cells and

‡ Present addresses: Department of General Physiology & Biophysics, The Panum Institute, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark (S.-P.O.). Department of Physiology, Mayo Pharmacology & Foundation, Rochester, Minnesota 55905, USA (D.E.C.).

suggested that they could be involved in the response to mechanical forces generated by blood flow. The channels were relatively non-selective and were opened by membrane stretching induced by suction. Here we report whole-cell patch-clamp recordings of single arterial endothelial cells exposed to controlled levels of laminar shear stress in capillary flow tubes. A K^+ selective, shear-stress-activated ionic current (designated $I_{K,S}$) was identified which is unlike previously described stretch-activated currents. $I_{K,S}$ varies in magnitude and duration as a function of shear stress (half-maximal effect at 0.70 dyn cm^{-2}), desensitizes slowly and recovers rapidly and fully on cessation of flow. $I_{K,S}$ activity represents the earliest and fastest stimulus-response coupling of haemodynamic forces to endothelial cells yet found. We suggest that localized flow-activated hyperpolarization of endothelium involving $I_{K,S}$ may participate in the regulation of vascular tone.

Endothelial cells were grown at low density on the inner surface of glass capillary tubes. Whole-cell patch-clamp recordings of single cells were made while flow through individual tubes was varied (Fig. 1a). Membrane potential changed when endothelial cells, grown as individual cells as well as in a continuous monolayer, were exposed to the shear stresses induced by laminar flow. At physiological ion concentrations (see legend to Fig. 1), the membrane-potential change was always in the hyperpolarizing direction and varied from 0 to 6 mV ($n=6$). Resting membrane potentials (V_m) were usually near the potassium equilibrium potential of -90 mV ($V_m = -77 \text{ mV}$, $n=26$). Under voltage clamp, flow induced an ionic current ($I_{K,S}$), which increased as a function of shear stress, τ (Fig. 1b). $I_{K,S}$ was measured as an inward current using conditions designed to maximize the size of the current (holding potential $V_h = -30 \text{ mV}$, symmetric 145 mM K^+). $I_{K,S}$ desensitized very slowly and rapidly returned to full levels when flow was stopped and restarted. τ (dyn cm^{-2}) was calculated for laminar flow in cylindrical tubes according to the equation

$$\tau = 4\mu Q / \pi r^3 \text{ (ref. 8)}$$

where μ represents fluid viscosity (0.0077 P), Q , the flow rate (ml s^{-1}) and r the internal radius of the tube ($0.0675 \pm 0.004 \text{ cm}$). The equation only approximates shear stress as there will be some slight acceleration of flow near the end of the tube due to the entering micropipette and boundary effects at the termination of the tube. The flow characteristic throughout these experiments was steady laminar (maximum Reynolds number, $R_e = 459$ calculated with $\tau = 16.5 \text{ dyn cm}^{-2}$, $Q = 0.5 \text{ ml s}^{-1}$). Flow rates varied between 0.16 and 31.0 ml min^{-1} to create laminar shear stresses of 0.08 – 16.5 dyn cm^{-2} . The relation between τ and $I_{K,S}$ was fitted by a simple sigmoid curve with half-maximal effect at 0.70 dyn cm^{-2} (Fig. 1b). The $I_{K,S}$ - τ relation saturated above 15 dyn cm^{-2} .

Figure 2a shows the current evoked by flow at various holding potentials. The current-voltage relation rectified sharply above the potassium holding potential (E_K) and displayed a negative slope in the depolarized range, similar to potassium inward-rectifier currents of other cell types⁹. When reversal potentials for $I_{K,S}$ were plotted as a function of extracellular K^+ , the relation had a slope close to that expected for an ion-selective membrane according to the Nernst relation (Fig. 2b). Like other inwardly-rectifying potassium currents, $I_{K,S}$ was blocked by extracellular perfusion with a solution containing 1 mM Ba^{2+} (Fig. 2c) and by internal and external caesium ($[Cs^+]_i = 145 \text{ mM}$; $[Cs^+]_o = 2 \text{ mM}$). Thus, $I_{K,S}$ is an inwardly-rectifying, potassium-selective current blocked by external barium.

$I_{K,S}$ was not an artefact due to flow-induced movement of the recording pipette; we were unable to reproduce the current by deliberate movements of the electrode during whole-cell recordings, or by sealing the electrode to a Sylgard ball and applying flow. Another potential artefact of high flow rates, washout of potassium from restricted extracellular spaces, was eliminated by the use of a high potassium concentration in the bulk extracellular solution. Accumulation under these conditions would vio-

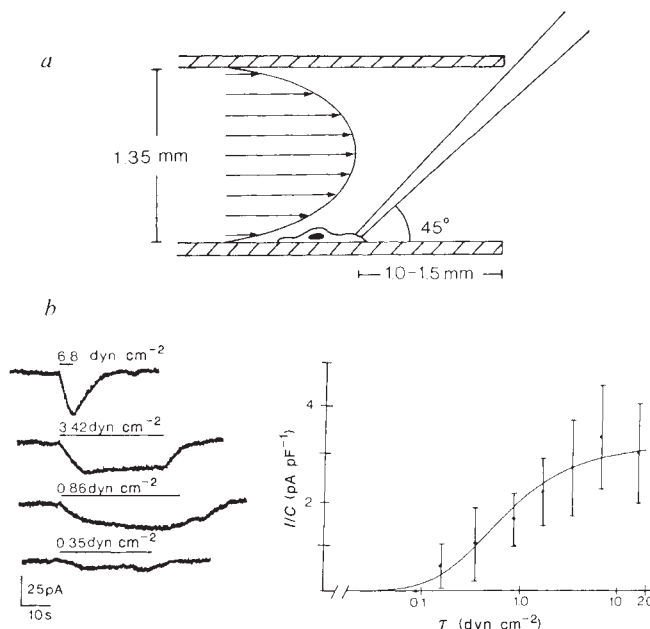


Fig. 1 a, Experimental design showing a patch-clamp pipette attached to an endothelial cell cultured in a laminar flow tube. The glass tube was perfused with the bathing medium using an infusion pump, and the flow profile is indicated at the site of the endothelial cell studied (enlarged).

Methods. Bovine aortic endothelial cells were isolated from calf aorta by collagenase digestion according to Schwartz²⁴ followed by subculture to passage 7–15. The endothelial cells were then brought into suspension by exposure to trypsin and EDTA, and the cell suspension was plated at subconfluent density by capillary action into glass tubes that had been precoated with 0.1% gelatin. Dulbecco's modified Eagle's medium contained 10 mM HEPES, 2 mmol ml^{-1} glutamine, 100 U ml^{-1} penicillin, $100 \mu\text{g ml}^{-1}$ streptomycin and was supplemented with 10% calf serum.

Cells were voltage-clamped and whole-cell currents were recorded using the patch-clamp technique²⁵. The signals were recorded with a List EPC-7 amplifier, stored on a digital recorder and displayed on a chart recorder. All curves were traced from the strip-chart recorder which filters with a cut-off frequency of $\sim 100 \text{ Hz}$. The tip resistance of the electrodes filled with the intracellular solution was 5 – $8 \text{ M}\Omega$. The ion composition of the extracellular solution was as follows (in mM): K^+ , 4 ; Na^+ , 150 ; Cl^- , 150 ; Ca^{2+} , 1 ; Mg^{2+} , 2 ; HEPES, 10 ; pH was adjusted to 7.4 with HCl. The high potassium intracellular solution contained (in mM): K^+ , 145 ; Cl^- , 116 ; Ca^{2+} , 1 ; Mg^{2+} , 2 ; EGTA, 11 ; HEPES, 10 ; GTP, 0.20 ; pH was adjusted to 7.2 with HCl. $T = 22 \pm 2^\circ\text{C}$. b, Relationship between fluid shear-stress amplitude at the cell surface and whole-cell current. The current was recorded in endothelial cells bathed in high-potassium solution and perfused with this solution at rates creating shear-stress forces of 0.086 – 16.5 dyn cm^{-2} . $V_h = -30 \text{ mV}$. Tracings showing the time course of whole-cell currents from one cell induced by four different flow rates. At 6.8 dyn cm^{-2} , longer pulse durations than shown here induced desensitization, producing an inaccurate I/C ($C = 15 \text{ pF}$). The whole-cell current normalized by cell capacitance is plotted as a function of the fluid shear stress (τ). The mean values $\pm$ s.d. of seven experiments are shown. The maximal current is $\sim 3.2 \text{ pA pF}^{-1}$ and half-maximal effect is induced by a shear stress of 0.70 dyn cm^{-2} .

late charge neutrality. The flow-induced current is specific to endothelial cells; heart cells exposed to flow under the same conditions as endothelial cells did not exhibit $I_{K,S}$.

We have previously described a muscarinic acetylcholine (ACh)-induced inward-rectifying K^+ current ($I_{K,ACh}$) in these cells¹⁰. As it has been reported that endothelial cells may produce ACh¹¹, we tested the hypothesis that flow could stimulate ACh release resulting in activation of the muscarinic-gated ion channel. As shown in Fig. 3, however, the ACh-activated current added to $I_{K,S}$ and was selectively blocked by atropine, whereas

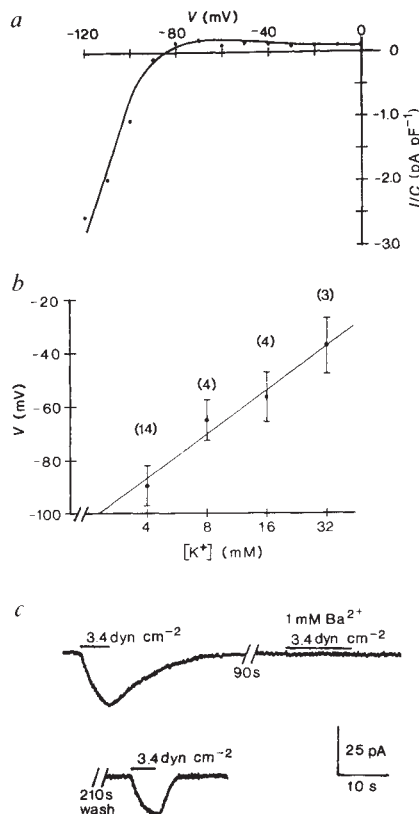


Fig. 2 *a*, Current-voltage curve of fluid shear-stress-induced current in an endothelial cell. The cell was bathed in a normal extracellular solution, ($[K^+]_o = 4$ mM), and the curve was obtained by holding the membrane potential at values ranging from -120 to 0 mV and exposing the cells to a fluid shear stress of 3.4 dyn cm^{-2} for ~ 15 s at each potential. The current showed inward rectification with a reversal potential at -90 mV and inward-going slope conductance of 91 pS pF^{-1} . *b*, Potassium selectivity of shear-stress induced current. $I-V$ relations were determined from endothelial cells bathed in and perfused with extracellular solution containing four different potassium concentrations (4, 8, 16 and 32 mM), obtained by replacing NaCl with KCl to the desired concentrations. The reversal potentials were determined and plotted as a function of the extracellular potassium concentration. Mean values $\pm$ s.d. are shown and the number of experiments in each group is indicated above the points. The line through the points is a linear least-square fit, with a slope of 55 mV per 10-fold increase in $[K^+]_o$. *c*, Blocking of I_{KS} by barium. The tube was perfused with normal extracellular solution, which induced an inward current of 28 pA ($V_h = -110$ mV). Addition of 1 mM Ba^{2+} to the perfusate solution reversibly blocked I_{KS} .

I_{KS} was unaffected by atropine. Thus, I_{KS} and I_{KACh} are two distinct ion currents.

We have failed to observe single I_{KS} channels in numerous outside-out patches. This could be due to very low channel density (consistent with the small whole-cell current) or may be because the patch area is too small to allow sufficient deformation of the membrane by flow.

These studies report a new link between haemodynamically-related laminar flow and a K^+ current in endothelial cells. I_{KS} hyperpolarizes endothelial cells by gating an inward-rectifying K^+ -selective conductance at low levels of shear stress. In arterioles, τ ranges between about 5 and 20 dyn cm^{-2} for vessels 20–40 μ m in diameter¹², well within the range of I_{KS} activation (0.2–17.0 dyn cm^{-2}). We suggest that arterial vessel relaxation in response to increased blood flow¹³ is moderated directly or indirectly by I_{KS} . Relaxation of underlying smooth muscle cells could occur by gap-junctional coupling to endothelial cells^{14–16}. Relaxation of smooth muscle by direct hyperpolarization (elec-

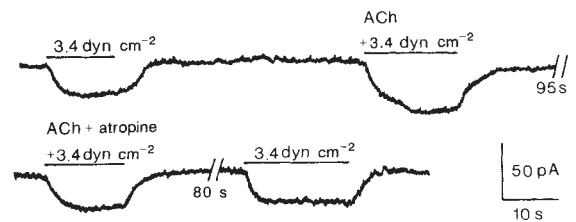


Fig. 3 Shear stress- and acetylcholine (ACh)-activated currents. Perfusion of the tube activated an inward current of 27 pA ($V_h = -30$ mV, 145 mM symmetrical $[K^+]$). When the cell was exposed to shear stress and 3 μ M ACh simultaneously, the inward current increased to 50 pA. When 3 μ M atropine blocked the ACh-induced current, the shear-stress-induced current alone was present. In a series of six cells mean ACh-induced current was 1.6 ± 0.6 pA pF^{-1} ($V_h = -30$ mV, 145 mM symmetrical $[K^+]$).

trode impalement) has been described for coronary arteries of the monkey¹⁷. Alternatively, hyperpolarization could modulate endothelial-derived relaxing factor (EDRF) release^{18–20} by unknown mechanisms.

In large arteries, high flow velocities (~ 40 cm s^{-1}) result in estimated average shear-stress levels of up to 30 dyn cm^{-2} with transient elevations as high as 100 dyn cm^{-2} (ref. 8), values outside of the activation range of I_{KS} . At locations adjacent to branches and flow dividers, however, there is a strong correlation between low mean shear stress (0 – 6 dyn cm^{-2}) and atherosclerotic lesions^{21,22}. The levels of shear stress in such regions fall within the activation range of I_{KS} .

The net I_{KS} has properties unlike the nonselective stretch-activated channels observed in endothelial cells⁷ and in skeletal muscle²³. The previously described endothelial single channel was permeable to K^+ , Na^+ , Ca^{2+} and Cs^+ , and thus would conduct an inward, depolarizing current under physiological conditions. An interesting possibility is that at low shear stress, I_{KS} hyperpolarizes endothelial cells, leading to dilation of arterioles or localized regions of large arteries, whereas high shear stress may gate the stretch-activated channels to cause the opposite effect, the response depending upon location of endothelial cell in the circulation and the channel type expressed.

This work was supported by the Danish MRC, NATO Science Fellowship Programme and the US National Heart, Lung and Blood Institute. We gratefully acknowledge discussions with Dr C. F. Dewey of MIT and our colleagues Drs B. Bean, R. Cotran, M. Gimbrone and P. Hess.

Received 26 August; accepted 6 November 1987.

1. Fry, D. L. *Circulation Res.* **22**, 165–197 (1968).
2. Flaherty, J. T. *et al. Circulation Res.* **30**, 23–32 (1972).
3. Dewey, C. F. Jr, Bussolari, S. R., Gimbrone, M. A. Jr & Davies, P. F. *J. biomech. Engng* **103**, 177–185 (1981).
4. Davies, P. F., Dewey, C. F. Jr, Bussolari, S. R., Gordon, E. J. & Gimbrone, M. A. *Jr J. clin. Invest.* **73**, 1121–1129 (1984).
5. Frangos, J. A., Eskin, S. G., McIntire, L. V. & Ives, C. L. *Science* **227**, 1477–1479 (1985).
6. Davies, P. F., Remuzzi, A., Gordon, E. J., Dewey, C. F. Jr & Gimbrone, M. A. *Jr Proc. natn. Acad. Sci. U.S.A.* **83**, 2114–2117 (1986).
7. Lansman, J. B., Hallam, T. J. & Rink, T. J. *Nature* **325**, 811–813 (1987).
8. Dewey, C. F. *Jr Adv. exp. Med. Biol.* **115**, 55–103 (1979).
9. Hille, B. *Ionic Channels in Excitable Membranes* 111–114 (Sinauer, Sunderland, Mass., 1984).
10. Olesen, S.-P., Davies, P. F. & Clapham, D. E. *Circulation Res.* (in the press).
11. Parnavelas, J. G., Kelly, W. & Burnstock, G. *Nature* **316**, 724–725 (1985).
12. Arfors, K.-E. & Bergquist, D. *Thromb. Res.* **4**, 447–461 (1974).
13. Holtz, J., Forstermann, U., Pohl, U., Giesler, M. & Bassenge, E. J. *Cardiovasc. Pharmac.* **6**, 1161–1169 (1984).
14. Rhodin, J. A. G. *J. Ultrastruct. Res.* **18**, 181–223 (1967).
15. Davies, P. F. *Lab. Invest.* **55**, 5–24 (1986).
16. Segal, S. S. & Duling, B. R. *Circulation Res.* **61**, 1120–1125 (1987).
17. Mekata, F. *J. Physiol., Lond.* **371**, 257–265 (1986).
18. Furchgott, R. F. & Zawadzki, J. V. *Nature* **288**, 373–376 (1980).
19. Vanhoutte, P. M., Rubanyi, G. M., Miller, V. M. & Houston, D. S. *A. Rev. Physiol.* **48**, 307–320 (1986).
20. Peach, M. J., Singer, M. A. & Loeb, A. L. *Biochem. Pharmac.* **34**, 1867–1874 (1985).
21. Caro, C. G., Fitzgerald, J. M. & Schroter, R. C. *Proc. R. Soc. B* **177**, 109–159 (1971).
22. Ku, D. N., Giddens, D. P., Zarins, C. K. & Glagov, S. *Arteriosclerosis* **5**, 293–301 (1985).
23. Sachs, F. in *Ionic Channels in Cells and Model Systems* (ed. Latorre, R.) 181–193 (Plenum, New York, 1986).
24. Schwartz, S. M. *In Vitro* **14**, 966–980 (1978).
25. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).

Cloning of the mycobacterial epitope recognized by T lymphocytes in adjuvant arthritis

Willem van Eden*, Jelle E. R. Thole†, Ruurd van der Zee‡, Alie Noordzij*, Jan D. A. van Embden†, Evert J. Hensen* & Irun R. Cohen§

* Department of Infectious Diseases and Immunology, Veterinary Faculty, State University of Utrecht, 3508 TD Utrecht, The Netherlands

† Laboratory of Bacteriology, National Institute of Public Health and Environmental Hygiene, Bilthoven, The Netherlands

‡ Laboratory for Medical Microbiology, State University of Groningen, Groningen, The Netherlands

§ Department of Cell Biology, Weizmann Institute of Science, Rehovot, Israel

Adjuvant arthritis (AA) is a chronic disease inducible in rats by immunization with an antigen of *Mycobacterium tuberculosis*¹. After the isolation of arthritogenic T-cell lines² and clones³, it became possible to demonstrate that the critical *M. tuberculosis* antigen contained an epitope cross-reactive with a self-antigen in joint cartilage⁴⁻⁶. Like AA rats, patients suffering from rheumatoid arthritis demonstrated specific T-lymphocyte reactivity to the *M. tuberculosis* fraction containing the cross-reactive epitope⁷. To characterize the critical *M. tuberculosis* epitope we used AA T-cell clones to screen mycobacterial antigens expressed in *Escherichia coli* and genetically engineered truncated proteins and synthetic peptides. The AA T-cell clones recognized an epitope formed by the amino acids at positions 180–188 in the sequence of a *Mycobacterium bovis* BCG antigen⁸. Administration of this antigen to rats induced resistance to subsequent attempts to produce AA.

Some rat strains develop an autoimmune-like arthritis after immunization with *M. tuberculosis* (Mt) (ref. 1), but AA can sometimes be induced without using Mt (ref. 9). To investigate the relation between AA and immunization with Mt, we raised lines of arthritogenic anti-Mt T cells from Lewis rats immunized to induce AA (ref. 2). Clone A2b of one such line produced arthritis in heavily irradiated (750R) Lewis rats³, and recognized both an Mt antigen and a fraction of cartilage proteoglycan

Table 1 Proliferative response of T-cell clones A2b and A2c to the 65K *M. bovis* BCG protein

	Mt	65K protein	<i>E. coli</i> control	Ovalbumin	Concanavalin A
A2b	180 (±21)	500 (±64)	2.9 (±0.4)	—	430 (±41)
A2c	304 (±18)	516 (±44)	1.5 (±0.2)	—	390 (±28)
C1a	—	1.5 (±0.1)	1.2 (±0.4)	45 (±5.1)	64 (±6.3)

Proliferative response of T-cell clones A2b, A2c and C1a in the presence of heat-killed *M. tuberculosis* H37Ra (Mt) (Difco), purified 65K *M. bovis* BCG recombinant protein⁸, and a control protein fraction (10 µg ml⁻¹) purified from non-recombinant *E. coli*, ovalbumin (20 µg ml⁻¹) and concanavalin A (2.5 µg ml⁻¹). Cloning and maintenance conditions of T cells have been described previously³. Proliferative responses were measured by 16 h [³H]thymidine incorporation in cells (2 × 10⁴ well) cultured for 4 d in the presence of irradiated (1,500 R) syngeneic thymocytes (2 × 10⁶ per well) as accessory cells. The counts per minute (c.p.m.) for [³H]thymidine incorporation were measured in triplicate test cultures and divided by the mean of triplicate cultures without antigen. The means of the resulting ratios are shown as stimulation index (SI, c.p.m. test per c.p.m. control without antigen) ± 1 standard deviation. Culture medium was DMEM (Gibco) supplemented with 1% rat serum, 5 × 10⁻⁵ M 2-mercaptoethanol, 2 mM glutamine, 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin. The 65K recombinant protein was prepared from the 65K over-producing *E. coli* host strain 1046, which carries the lambda repressor CI857-encoding plasmid pCI857 and the 65K protein-encoding plasmid pRIB1000 (ref. 12). The control *E. coli* K12 was obtained from strain 1046, carrying pCI857 and the cloning vector pPLc 236 (ref. 12).

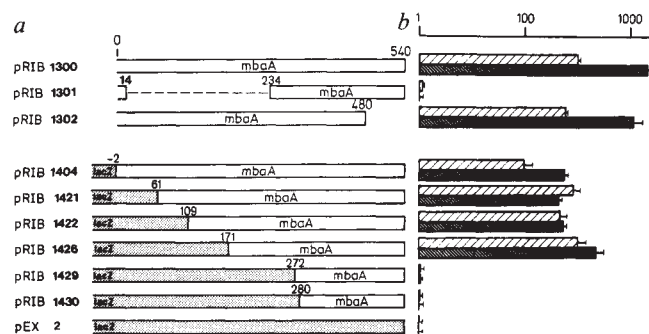
probably associated with the core protein⁴. An acetone-precipitable fraction of Mt containing the epitope that is cross-reactive with cartilage was recognized not only by AA T lymphocytes, but also by T lymphocytes from patients with rheumatoid arthritis⁷. Another clone, A2c, functioned as a suppressor-inducer T cell^{5,10}. Rats treated with A2c acquired resistance to AA (ref. 5).

We have used clones A2b and A2c to probe individual mycobacterial antigens expressed in *E. coli* recombinants. Table 1 shows the responses of anti-Mt clones A2b and A2c, and of control clone C1a, responsive to ovalbumin. It can be seen that both clones A2b and A2c responded not only to whole Mt, but also to an *M. bovis* BCG antigen of relative molecular mass (*M_r*) ~65,000 (65K) and expressed by *E. coli*. A control protein

Fig. 1 *a*, Map of deletion mutants of the gene encoding the 65K protein fused to *lacZ* and truncated derivatives of *mbaA* expressed in the plasmids indicated. *b*, Proliferation of T lymphocyte clones A2b (upper, light shaded bars) and A2c (lower, dark shaded bars) in response to the *M. bovis* BCG 65K protein (*mbaA*), truncated derivatives and fusion proteins with β -galactosidase.

Methods. Plasmid pRIB1300 is an *EcoRI*-*Bst*MII deletion mutant of plasmid pRIB1011 (ref. 8). Plasmid pRIB1301 is an in-frame *XhoI* deletion derivative of pRIB1300, and pRIB1302 is an *AccI*-*SalI* deletion mutant of pRIB1300. Plasmid pRIB1300 expressed the native *mbaA* protein composed of 540 amino-acid residues. The deletion mutants pRIB1301 and pRIB1302 express truncated proteins. In the plasmids pRIB1404, pRIB1421, pRIB1422, pRIB1426, pRIB1429 and pRIB1430, the carboxy-terminal part of *lacZ* is fused with parts of the 65K protein-coding gene¹⁹. The number of the residue from the amino-terminal of *mbaA* is indicated for the coding gene fused to *lacZ*. In pRIB1404, *lacZ* is fused to the *M. bovis* DNA 7 base pairs upstream of the *mbaA* start codon and therefore the fusion protein contains two extra residues. All *lacZ* fusions were constructed using the *lacZ*-containing expression vector pEX2 (ref. 20).

Proliferative responses of T-cell clones A2b and A2c in the presence of various mutant proteins were measured as in Table 1. Mean stimulation indexes ± 1 standard deviation (SI ± 1 s.d.) are given. Antigens were prepared as follows. *E. coli* K12 M1070 (ref. 8) carrying the plasmids pRIB1300, pRIB1301 or pRIB1302 was used to produce the 65K protein or truncated derivatives. Induced cultures were resuspended in 100 mM Tris, 5 mM EDTA, pH 8 and sonicated by 8 pulses of 30 s at 80 Watts. The resulting lysates contained the 65K protein up to 30% (pRIB1300), or truncated derivatives up to 5% (pRIB1301 or pRIB1302) of the total protein content. Final total protein concentration used during T-cell stimulation was 1 µg ml⁻¹. *E. coli* K12 M1070 carrying the plasmid pEX2 or derivatives was used to induce the various β -galactosidase fusion proteins²⁰. Induced cells were resuspended in 100 mM Tris, 10 mM EDTA, pH 8.0 containing lysozyme 0.1 mg ml⁻¹, freeze-thawed once and ultrasonicated by three pulses of 30 s at 70 Watts. About 40% of the total protein in the resulting lysates consisted of fusion protein. Final total protein concentration used during T-cell stimulation was 0.5 µg ml⁻¹.



Peptide	Sequence	T-cell response SI ($\pm$ s.d.)		
		A2b	A2c	Z1a
	171 181 191 GVITVEESNTFGLQLELTEGMRFDKGYISG			
153-171	<-----+	<1	=1	ND
174-192	+-----+	16 ($\pm$ 2)	11 ($\pm$ 2)	<1
180-196	+-----+	33 ($\pm$ 5)	120 ($\pm$ 8)	<1
180-188	+-----+	47 ($\pm$ 4)	58 ($\pm$ 3)	<1
183-196	+-----+	9.2 ($\pm$ 0.3)	2.9 ($\pm$ 0.8)	<1
185-196	+-----+	<1	<1	ND
190-200	+-----+	<1	<1	ND
197-218	+-----<	<1	<1	=1
Mt		180 ($\pm$ 21)	304 ($\pm$ 18)	<1
BP		=1	<1	162 ($\pm$ 18)

Fig. 2 T-cell epitope mapping with synthetic peptides of the 65K *M. bovis* BCG protein. The amino-acid sequence of the protein is shown for the area critical for T-cell recognition by arthritis clones A2b and A2c. The amino-acid sequence is predicted from the nucleotide sequence^{12,21}. The underlined and bold-printed sequence is the region essential for T-cell recognition.

Methods. Peptides 180-196, 180-188, 183-196, 185-196 and 190-200 were prepared by solid-phase techniques²² using a Labortec SP640 peptide synthesizer. Boc amino acids were coupled as the preformed symmetric anhydrides (hydroxybenzotriazole esters for arginine, asparagine and glutamine). *N,N*-Diisopropylcarbodiimide was the activating agent. Peptides were deprotected and cleaved from the resin with trifluoromethanesulphonic acid in TCA in the presence of thioanisole. After separation from the resin by filtration, the peptides were precipitated and washed with dry ether. Finally, peptides were desalted on G10 Sephadex in 5% acetic acid and lyophilized. Peptides 153-171 (ref. 23) and 174-192 and 197-218 were gifts. T-cell responses were determined in the standard lymphocyte proliferative assay described in Table 1. Mean SI $\pm$ s.d. are given. Data are shown for responses at a peptide concentration of 1 μ g ml⁻¹. BP (basic protein of myelin) and Mt were used at 10 μ g ml⁻¹. ND, not determined.

fraction from *E. coli* not transfected with mycobacterial DNA did not stimulate A2b or A2c. Control clone C1a responded to ovalbumin but not to any of the Mt or *E. coli* antigens. All three clones responded to the mitogen concanavalin A. Clones A2b and A2c did not respond to other cloned mycobacterial antigens of 12K, 18K, 28K or 34K (not shown)¹¹.

To identify the epitopes more precisely, we tested the responses of clones A2b and A2c to fragments of the 65K protein which were obtained either from deletion mutants of the gene for the protein, or from deletion mutants of the gene after fusion with the β -galactosidase gene or by synthesis of peptide regions included in the 65K protein. Figure 1 shows a map of the deletion mutants, the parts of the 65K coding gene fused to *lacZ*, and the stimulation of clones A2b and A2c by the expressed proteins. It can be seen that both clones responded to protein fragments that lacked the amino acids up to residue 171 (see pRIB1426) but deletions extending to amino acid 234 (pRIB1301) showed no antigenic activity, indicating that the epitopes were probably located between amino-acid residues 171-234. A deletion of the 60 carboxy-terminal amino acids (pRIB1302) did not abolish activity.

Synthetic peptides based on the sequence of the 65K *M. bovis* BCG protein were tested for stimulation of both T-cell clones. Figure 2 shows the amino-acid sequence of the 65K *M. bovis* BCG protein¹² from residue 171-200 and of the informative synthetic peptides. Note that both clones A2b and A2c responded to peptides 174-192, 180-196 and 180-188, whereas Z1a, the control clone with a specificity for myelin, did not. The responses of A2b and A2c to peptide 183-196 were significantly lower than to 180-196 and 180-188, and were negative in the case of 185-196. Therefore, the critical epitope for both clones resides in the 180-188 sequence.

An amino acid-homology search showed some similarity between the 180-188 sequence and the link protein of rat proteoglycan¹³, with four of the nine amino acids identical. No similarity was observed with published sequences of the core protein of chicken and rat proteoglycan^{14,15}.

Thus arthritogenic clone A2b and protective clone A2c both recognize a single nonapeptide epitope of the 65K protein of mycobacteria. It was important to establish whether active immunization with the 65K antigen would activate A2b-like T lymphocytes and induce AA, or if it would activate A2c-like T suppressor-inducers and induce resistance to AA. Unlike immunization with whole mycobacteria, the administration of

the 65K antigen emulsified in oil did not induce AA (Table 2), but the immunized rats did show resistance to a subsequent attempt to induce AA by immunization to whole Mt in oil (Table 2). Thus the engineered 65K antigen, although recognizable *in vitro* by the arthritogenic T lymphocyte clone A2b, induced resistance to AA, presumably by favouring the emergence of A2c-like (suppressor-inducer) T cells. Indeed, A2b cells treated to cause aggregation of membrane components were found, similarly to A2c cells, to induce anti-idiotypic immunity and remission of established AA (ref 16).

A study of a small number of patients with rheumatoid arthritis showed that T-cell recognition of the 65K antigen correlated with T-cell recognition of the acetone-precipitable fraction of whole Mt (ref. 7 and R. R. P. de Vries, personal communication). Therefore, the 65K antigen could be a specific target for T-lymphocyte recognition in such patients. A possible function for the 65K protein has been suggested¹⁷ after extensive homology was found with a sequence encoded by an essential *E. coli* gene involved in messenger processing. It appears that the 65K antigen is a heat-shock protein (D. B. Young, personal communication), related to homologous proteins in higher eukaryotes as well as in bacteria. Self-epitopes present in stress proteins could be a necessary by-product of their phylogenetic conservation. Antigens cross reactive with the 65K *M. bovis* BCG protein are present in some mycobacteria⁸ and this antigen is also related to another shared by more than 50 different

Table 2 65K protein of *M. bovis* BCG protects against AA

Immunizing agent	Primary immunization Arthritis incidence	Secondary AA challenge (after 35 d) with Mt in oil	
		Arthritis incidence	Clinical grade
Saline	0/8	8/8	Severe
65K protein	0/8	1/8	Very mild
<i>E. coli</i> control protein	0/7	6/7	Severe

Groups of 7 or 8 Lewis rats were treated by intraperitoneal inoculation of saline, 65K protein (50 μ g) or *E. coli* control protein (weight equivalent to *E. coli* content of 50 μ g of 65K protein) in oil. After 35 d the susceptibility to induction of AA was tested by challenging the rats intracutaneously at the base of the tail with 1 mg heat-killed Mt in oil. Incidence of arthritis was checked by daily inspection of the rat joints and confirmed by histological examination.

bacterial species, including *Klebsiella*, *Shigella*, *Salmonella*, *Yersinia* and *Campylobacter*, all suspected of being involved in human arthritis¹⁸. Possibly humans could be exposed to the 180–188 epitope, or to a cross-reactive epitope associated with different environmental bacteria under conditions which might influence their susceptibility to autoimmune arthritis. If the 180–188 epitope is involved in the pathophysiology of arthritis, a purified peptide containing the epitope might induce therapeutic suppression of the disease process in the same way

as the 65K antigen in AA.

We thank Drs D. B. Young and T. M. Buchanan for synthetic peptides, W. J. Keulen and P. Hermans for technical assistance, and Dr E. J. M. Hogervorst for support. J.E.R.T. received financial support from the World Health Organization Program for Vaccine Development. T-cell clones A2b and A2c were generated with the support of the NIH and the Crown Foundation. I.R.C. holds the Mauerberger Professorial Chair in Immunology.

Received 10 September; accepted 16 November 1987.

- Pearson, C. M. *Arthritis Rheum.* **7**, 80–86 (1964).
- Holoshitz, J., Naparstek, Y., Ben-Nun, A. & Cohen, I. R. *Science* **219**, 56–58 (1983).
- Holoshitz, J., Matitiai, A. & Cohen, I. R. *J. clin. Invest.* **73**, 211–215 (1984).
- van Eden, W. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5117–5120 (1985).
- Cohen, I. R., Holoshitz, J., van Eden, W. & Frenkel, A. *Arthritis Rheum.* **29**, 841–845 (1985).
- van Eden, W., Holoshitz, J. & Cohen, I. R. in *Autoimmunoregulation and autoimmune disease* (ed. Cruse, J. M. & Lewis, R. E.) 144–170 (Karger, Basel, 1987).
- Holoshitz, *et al. Lancet* **ii**, 305–309 (1986).
- Thole, J. E. R. *et al. Infect. Immunity* **50**, 800–806 (1985).
- Chang, Y. H., Pearson, C. M. & Abe, C. *Arthritis Rheum.* **23**, 62–71 (1980).
- Cohen, I. R. *Immun. Rev.* **94**, 5–12 (1986).
- Young, R. A. *et al. Nature* **316**, 450–452 (1985).
- Thole, J. E. R. *et al. Infect. Immunity* **55**, 1466–1475 (1987).

- Neame, T. J., Chrestner, J. E. & Baker, J. R. *J. biol. Chem.* **261**, 3519–3535 (1986).
- Sai, S., Tanaka, T., Kosher, R. A. & Tanzer, M. L. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5081–5085 (1986).
- Doegge, K., Fernandez, P., Hassell, J. R., Sasaki, M. & Yamada, Y. *J. biol. Chem.* **261**, 8108–8111 (1986).
- Lider, O., Karin, N., Shinitzky, M. & Cohen, I. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4577–4580 (1987).
- Young, D. B., Ivanyi, J., Cox, J. H. & Lamb, J. R. *Immun. Today* **8**, 215–219 (1987).
- Thole, J. E. R. *et al. Microbial Pathogenesis* (in the press).
- Thole, J. E. R. *et al. Infect. Immunity* (in the press).
- Stanley, K. K. & Luzio, J. P. *EMBO J.* **3**, 1429–1434 (1984).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Steward, J. M. & Young, J. D. *Solid Phase Peptide Synthesis* (Pearce Chemical Co., Illinois, 1984).
- Lamb, J. R. *et al. EMBO J.* **6**, 1245–1249 (1987).

Basic fibroblast growth factor fused to a signal peptide transforms cells

S. Rogelj, R. A. Weinberg, P. Fanning* & M. Klagsbrun*

Whitehead Institute, Cambridge, Massachusetts 02142, USA

* Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

Basic fibroblast growth factor (bFGF) is a potent growth and angiogenic factor that is found in abundance in tissues such as brain, hypothalamus, kidney and cartilage^{1,2}. Despite this copious production of bFGF, most of these tissues are not undergoing either active growth or angiogenesis, suggesting that bFGF activity must be regulated so as to prevent autostimulation of cell growth. In cultured cells, bFGF is associated mainly with cells and basement membranes and is not released into the medium^{3,4}. Prevention of release could be a mechanism for regulation of bFGF activity and may be a consequence of the apparent absence of a secretory-signal sequence in the bFGF protein⁵. Here we investigate whether this regulation can be overridden through the forced secretion of bFGF. Such secretion might provide the bFGF access to its receptor and in turn lead to autocrine transformation of the cell. We report that bFGF, as specified by a recombinant plasmid, is itself unable to induce such transformation, but acquires this ability after fusion with a secretory-signal sequence. The resulting transformants undergo unusual morphological alteration and display tumorigenicity.

We inserted a complementary DNA encoding bovine bFGF (ref. 5) into the pJ33 plasmid (J. Morgenstern, unpublished) expression vector (pbFGF, Fig. 1a). In a second clone, sequences specifying an amino-terminal immunoglobulin signal peptide of 19 amino acids, which mediates co-translational insertion of a nascent protein into endoplasmic reticulum, were fused to the bFGF cDNA (refs 6, 7). The fusion joins this signal peptide to the second amino-acid residue of the native bFGF (pIgbFGF; Fig. 1b). The amino terminus of the predicted primary translation product is illustrated in Fig. 1c. After translation this chimaeric protein should be inserted into the endoplasmic reticulum and the signal peptide cleaved off, resulting in a protein of relative molecular mass (M_r) 18,000 (18K) which is virtually indistinguishable in size from normal bFGF.

To ascertain the transforming potential of these expression clones, the plasmids were transfected into NIH 3T3 cells⁸. This cell line is known to express bFGF receptors and respond

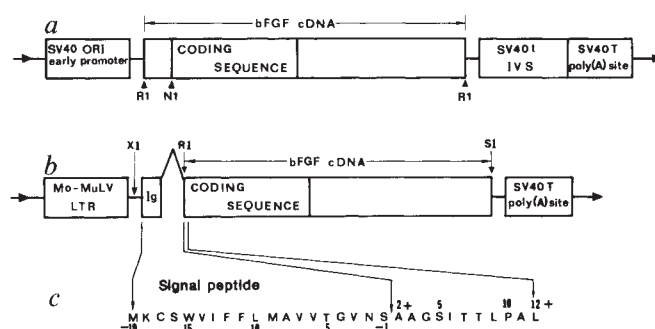


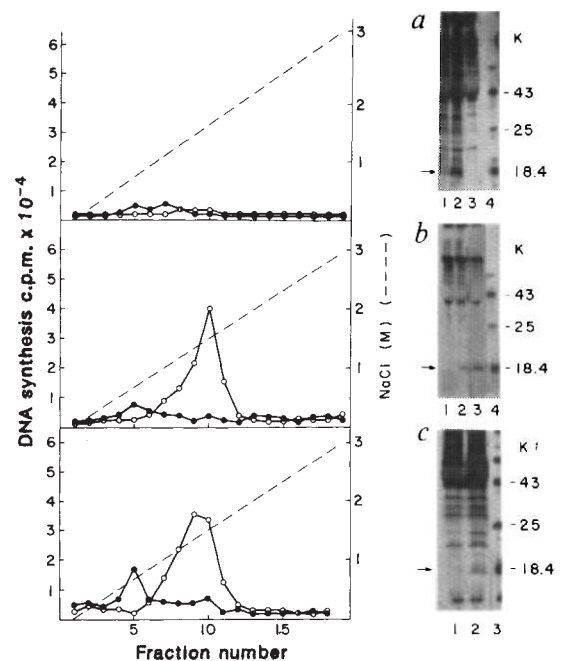
Fig. 1 Basic FGF expression constructs (R1, *EcoRI*; N1, *NcoI*; X1, *XbaI*; S1, *SalI* restriction site). a, pbFGF plasmid: the SV40 early promoter drives constitutive expression of the native bovine brain bFGF. b, pIgbFGF plasmid: a murine leukaemia virus long terminal repeat (LTR) containing promoter/enhancer sequences drives constitutive expression of the immunoglobulin signal peptide-bFGF fusion protein. c, Amino terminus of the predicted primary translation product of the chimaeric IgbFGF protein. The 19-amino-acid signal peptide is fused to the second amino acid of the native bovine brain bFGF.

Methods. The pbFGF plasmid was constructed by inserting the *EcoRI*-flanked bFGF cDNA sequence from pJ11-1 plasmid⁵ into the 5' to 3' orientation behind the SV40 early promoter sequences into the *EcoRI* site of the pJ33 mammalian expression vector. The pIgbFGF expression plasmid was constructed as follows. The pJ11-1 plasmid was digested with *NcoI* restriction enzyme to remove all of the 5'-end and some of the 3'-end non-coding sequences. The resulting 1,040 base pair (bp) cDNA fragment was blunted with mung-bean nuclease and ligated to 8-mer *EcoRI* linkers. After digestion with *EcoRI*, the fragment was redigested with *SspI* to further remove ~430 bp of 3'-end non-coding sequence. The isolated 606-bp fragment was first ligated to *SalI* linkers and then digested with *SalI* restriction enzyme. The purified 610-bp fragment was ligated into the pUCDS3 vector⁸ predigested with *EcoRI* and *SalI* and with the synthetic EGF sequences removed.

mitotically to bFGF treatment (refs 3, 4 and unpublished observations). The pbFGF plasmid did not induce any foci of transformants, even though these cultures were observed for four weeks. A control plasmid (pUCDS5) containing only the immunoglobulin signal sequence⁷ was also unable to induce focus formation. The pIgbFGF plasmid, however, induced foci with distinctive morphology. These foci were visible to the naked eye within 10 days of transfection, but were present only at a low frequency of ~40 foci per 8 μ g DNA per 10⁶ cells; but after two more weeks in culture, a second wave of foci appeared that

Fig. 2 Analysis of bFGF biosynthesis in monoclonal cell lines expressing bFGF and Ig bFGF by heparin-affinity chromatography and immunoprecipitation. The three panels on the left show heparin-affinity chromatography elution profiles. Open circles represent cell lysates, closed circles represent conditioned media. The three panels on the right show the immunoprecipitations of cell lysates. Arrows show the positions of bFGF. *a*, NIH-NM2 control cells. Lane 1: NIH-NM2; lane 2: NIH-BNM7; lane 3: NIH-BNM1 (another NIH 3T3 neomycin-resistant control cell line); lane 4: molecular weight markers. *b*, NIH-BNM7 cell line expressing bFGF. Lane 1: peptide control (immunoprecipitate of NIH-BNM7 in presence of excess peptide representing amino acids 33–43 of bFGF); lane 2: NIH-BNM7; lane 3: bovine aortic endothelial cells (BAE) (ref. 4); lane 4: molecular weight markers. *c*, NIH-IgBNM6-1 cell line expressing Ig bFGF. Lane 1: NIH-IgBNM6-1 peptide control (this cell line immunoprecipitated in the presence of excess peptide as above); lane 2: NIH-IgBNM6-1; lane 3: molecular weight markers.

Methods. NIH 3T3 cells transfected with pSV2-neo alone (NIH-NM2), with pbFGF and pSV2-neo (NIH-BNM7), and with pIg bFGF and pSV2-neo (NIH-IgBNM6-1) were grown in culture and bFGF synthesis was assayed by immunoprecipitation and by heparin-Sepharose affinity chromatography. NIH-NM2, NIH-BNM7, and NIH-IgBNM6-1 cells produced 0.04 ± 0.02 , 0.73 ± 0.23 , 0.8 ± 0.35 units of cell-associated growth factor activity per 10^4 cells respectively, compared to 0.89 ± 0.26 units per 10^4 cells synthesized by bovine endothelial cells⁴. Immunoprecipitation of bFGF was carried out as follows. Cells were grown in DMEM (Gibco) supplemented with 10% calf serum, penicillin and streptomycin to a density of $\sim 3 \times 10^6$ cells per 100 mm dish. The cells were radiolabelled for 16 h using 5 ml of cysteine-free DMEM media (Gibco) supplemented with 5% calf serum, penicillin and streptomycin and 50 μ Ci of [35 S]cysteine (specific activity $> 1,000$ Ci mmol⁻¹, New England Nuclear) per 100 mm dish. After removal of media, the cells were washed twice with PBS and scraped into 1 ml cell lysis buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM MgCl₂ and 0.5% (v/v) Nonidet P40 detergent). Following careful resuspension and a 5 min incubation on ice, the nuclei were pelleted with a 5 min centrifugation in an Eppendorf microfuge. The supernatant was pre-adsorbed with 5 μ l normal rabbit serum and 40 μ l protein A-Sepharose for 4 h at 4°C. The protein A-Sepharose was removed by a 1 min microfuge centrifugation. The supernatant was incubated for 16 h at 4°C with 5 μ l of anti-amino-terminus peptide rabbit antibody^{3,4}. As a control, 5 μ g of immunizing peptide were also added to an identical sample. Twenty μ l protein A-Sepharose was added for one hour of further incubation. The pellet was washed four times with cell lysis buffer with 1% Triton X-100, 1% sodium deoxycholate and 1% aprotinin solution (Sigma). Twenty μ l SDS-PAGE sample buffer was added to the washed pellet and the sample electrophoresed on SDS-PAGE (ref. 22). Heparin-Sepharose affinity chromatography was carried out as described^{3,4}. Cells were grown for 5–6 days until they reached confluence. Conditioned medium (100 ml) was removed, and cells ($0.5\text{--}1.5 \times 10^8$) were resuspended at 10^7 cells per ml in 10 mM Tris-HCl, pH 7, 1 M NaCl, and lysed by sonication. Cell lysates diluted to 0.1 M NaCl with conditioned medium were applied to individual heparin-Sepharose columns (3 ml) and bFGF was eluted with an 80 ml gradient of 0.1 M–3 M NaCl gradient at a flow rate of 30–40 ml h⁻¹. Fractions (4 ml) were collected and tested for ability to stimulate DNA synthesis in BALB-c/3T3 cells^{3,4}.



was 20-fold greater in number. These delayed foci did not arise through seeding of cells from the foci formed initially because they were also seen when the monolayer was overlaid with soft agar immediately after transfection.

Clonal cell lines synthesizing bFGF and Ig bFGF proteins were derived by co-transfecting the expression vectors with the dominant selectable marker pSV2-neo (ref. 9). Toxicity of the pIg bFGF clone in these cotransfections was measurable; addition of 8 μ g pIg bFGF to 0.1 μ g transfected pSV2-neo DNA resulted in a tenfold reduction of G418-resistant colonies. A similar inhibitory effect has been described for the Abelson murine leukaemia virus transforming gene¹⁰. Possibly this could account for the extremely low focus-forming efficiency and the dramatic reduction in the number of neomycin-resistant colonies appearing after co-transfection with pIg bFGF.

Neomycin-resistant colonies were expanded into cell lines and screened for bFGF activity increased over control NIH 3T3 cells transfected with pSV2-neo alone. Figure 2 shows the analyses of three selected cell lines: a neomycin-resistant control NIH 3T3 cell line, NIH-NM2; a bFGF-expressing cell line, NIH-BNM7; and an Ig bFGF-expressing cell line, NIH-IgBNM6-1. Basic FGF biosynthesis in these cell lines was analysed by immunoprecipitation of cell lysates and by heparin affinity chromatography of both cell lysates and conditioned medium^{3,4}.

Control NIH 3T3 cells did not synthesize detectable amounts of bFGF (Fig. 2a), but both NIH-BNM7 cells (Fig. 2b) and NIH-IgBNM6-1 cells (Fig. 2c) synthesized immunoprecipitable bFGF (Fig. 2, right) and biologically active, cell-associated bFGF that eluted from heparin-Sepharose columns at 1.5 M NaCl. The immunoprecipitable bFGF could be displaced

by an excess of the immunizing peptide used to induce anti-bFGF antibody. This material co-migrated with endothelial bFGF of $M_r \sim 18$ K (ref. 4), indicating that the Ig bFGF signal peptide was indeed cleaved off (Fig. 2c). The additional proteins of ~ 19 K and 23K immunoprecipitated from the NIH-IgBNM6-1 cells could represent post-translationally modified bFGF, suggesting passage through the endoplasmic reticulum and the Golgi. In both cell types, bFGF was found associated with the cells rather than with the conditioned medium. No immunoprecipitable or heparin-binding bFGF was found in the conditioned medium of NIH-IgBNM6-1 cells. Culture of bovine aortic endothelial cells in the presence of the NIH-IgBNM6-1 conditioned media, or even co-culture of these two cell types, did not result in an accelerated proliferation of the endothelial cells.

The growth patterns of NIH-NM2, NIH-BNM7 and NIH-IgBNM6-1 cells were compared (Fig. 3). NIH-BNM7 cells (Fig. 3b) grew minimally in soft agar. Their morphology was not very different from the control cells (Fig. 3a), although some changes were noted: the growth rate in 2% serum was roughly doubled, and the saturation density was about fourfold that of NIH-NM2 control cells.

The morphology of NIH-IgBNM6-1 cells differed dramatically from both the control NIH-NM2 cells and the NIH-BNM7 cells and was identical to cells in pIg bFGF-induced foci (Fig. 3c). The NIH-IgBNM6-1 cells grew as large stellate aggregates that barely adhered to the dish. Despite their active growth in semisuspension in media, growth in soft agar was minimal.

Cell lines expressing Ig bFGF were highly tumorigenic in syngeneic NIH/NSF mice, producing rapidly growing tumours within a week in all 50 mice tested. In contrast, no tumours

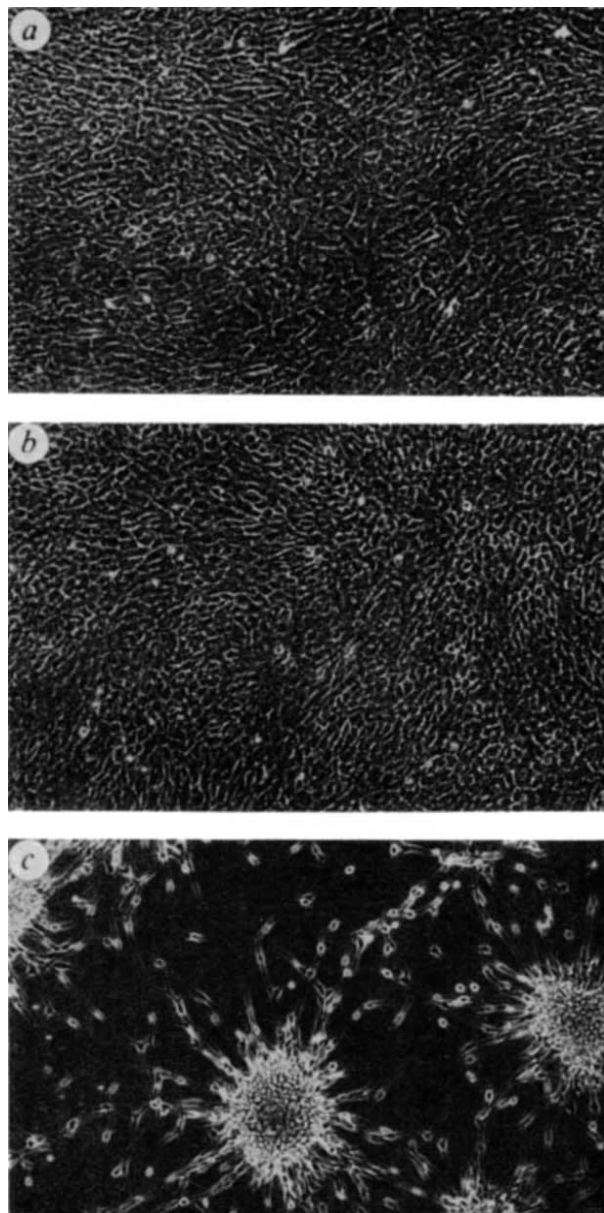


Fig. 3 Photomicrographs of neomycin-resistant monoclonal cell lines of NIH 3T3 cells transfected with *a*, pSV2-neo alone (NIH-NM2); *b*, pSV2-neo with pbFGF (NIH-BNM7); *c*, pSV2-neo with plgbFGF (NIH-IgBNM6-1). Cells were grown in 100 mm Costar dishes and photographed with technical PAN 2415 film using a Nikon inverted microscope.

were induced by control NIH 3T3 cells and only one tumour in 27 mice was induced by cells expressing bFGF. The NIH-IgBNM tumours could be cultured in the presence of neomycin and shown to produce substantial amounts of immunoprecipitable, heparin-binding, biologically active bFGF (data not shown), demonstrating that the tumours were derived from bFGF-producing cells.

More than 40 neomycin-resistant cell lines were also obtained by co-transfection of the selectable marker with plgbFGF. Many of these expressed barely detectable bFGF activity and did not display altered morphology at the time of isolation. However, during the first few rounds of passage, foci with stable transformed morphology appeared in cultures grown from these lines and expressed more bFGF than the parental clonal line when expanded. This delayed focus formation is reproducible and occurs in 90% of the neomycin-resistant cell lines co-transfected with plgbFGF.

We cannot explain the delayed appearance of foci of transformants which occurs at a low, but predictable, frequency in virtually all cell lines that have acquired pIgbFGF DNA. This event may represent a change in the expression of pIgbFGF DNA in these cells (for example, de-repression of the Mo-MuLV LTR promoter activity due to its demethylation), or may reflect a host-cell change that confers responsiveness on the acquired gene.

These results show that cells expressing comparable amounts of bFGF have a phenotype that depends on whether or not this protein has a signal peptide to allow entrance into the endoplasmic reticulum. Basic FGF lacking a signal peptide, as found in endothelial cells in culture, in tissues *in vivo*, and in the NIH-BNM cell lines, does not appear to trigger autocrine transformation of cells. In contrast, addition of the signal peptide to this growth factor has a profound effect on cellular phenotype and tumorigenic potential. The cell lines expressing IgbFGF are transformed and tumorigenic even though no bFGF is apparent in the medium. Since IgbFGF is found as a processed 18K protein, it has presumably undergone cleavage by signal peptidases responsible for the removal of signal sequence in the endoplasmic reticulum. It could be subsequently secreted and then immediately sequestered at the cell surface, activating a mitogenic pathway and completing the autocrine loop through an extracellular step. An alternative hypothesis put forward to explain transformation by the *sis* oncogene^{11,12,13} is that the signal peptide associated with bFGF directs this protein into the same cellular compartments that are encountered by bFGF receptor molecules during their own post-translational maturation. This growth factor may then bind its receptor at an intracellular location. A mitogenic response and resulting autocrine transformation could therefore be triggered even before the complex reaches the plasma membrane.

The present results show that the gene for bFGF, like those of other growth factors¹⁴⁻¹⁸, is a potential oncogene. Recent studies have shown that three different oncogenes encode FGF-homologous proteins: *int-2*, *hst*, and a bladder oncogene¹⁹⁻²¹. Whether these oncogenes produce biologically active, secreted bFGF-like growth factors has not yet been determined. The amino-termini of the proteins encoded by these oncogenes contain hydrophobic sequences that may serve as signal peptides. Cell lines expressing the chimaeric signal peptide bFGF could be used to investigate how bFGF and homologous genes act as oncogenes.

We wish to thank Drs Judith Abraham and John C. Fiddes of California Biotechnology Inc. for the bFGF cDNA clone, David Stern for the pUCDS3 and pUCDS5 plasmids, Shelly Bernstein for help and advice in tumorigenicity studies, and Yossi Yarden for advice and support.

Received 9 September; accepted 12 November 1987.

1. Folkman, J. & Klagsbrun, M. *Science* **235**, 442-447 (1987).
2. Baird, A. *et al. Recent Progress in Hormone Research* **42**, 143-205 (1986).
3. Klagsbrun, M., Sasse, J., Sullivan, R. & Smith, J. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2448-2452 (1986).
4. Vlodavsky, I. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 2292-2296 (1987).
5. Abraham, J. A. *et al. Science* **233**, 545-548 (1986).
6. Loh, R. Y., Bothwell, A. L. M., White-Scharf, M. E., Imanishi-Kari, T. & Baltimore, D. *Cell* **33**, 85-93 (1983).
7. Stern, D. F., Hare, D. L., Cecchini, M. A. & Weinberg, R. A. *Science* **235**, 321-324 (1987).
8. Smolkin, D., Gianni, A. M., Rozenblatt, S. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4910-4913 (1975).
9. Southern, P. J. & Berg, P. *J. molec. appl. Genet.* **1**, 327-332 (1982).
10. Ziegler, S. F., Whitlock, C. A., Goff, S. P., Gifford, A. & Witte, O. N. *Cell* **27**, 477-486 (1981).
11. Leal, F., Williams, L. T., Robbins, K. C. & Aaronson, S. A. *Science* **230**, 327-330 (1985).
12. Robbins, K. C., Leal, F., Pierce, J. H. & Aaronson, S. A. *EMBO J.* **4**, 1783-1792 (1985).
13. Betsholtz, C., Westermark, B., Ek, B. & Heldin, C.-H. *Cell* **39**, 447-457 (1984).
14. Lang, R. A., Metcalf, D., Gough, N. M., Dunn, A. R. & Gonda, T. D. *Cell* **43**, 531-542 (1985).
15. Rettenmier, C. W. *et al. Molec. cell. Biol.* **7**, 2378-2387 (1987).
16. Watanabe, S., Lazar, E. & Sporn, M. B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1258-1262 (1987).
17. Rosenthal, A., Lindquist, P. B., Bringman, T. S., Goeddel, D. V. & Derynck, R. *Cell* **46**, 301-309 (1986).
18. Finzi, E. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 3733-3737 (1987).
19. Dickson, C. & Peters, G. *Nature* **326**, 833 (1987).
20. Taira, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 2980-2984 (1987).
21. Marx, J. L. *Science* **237**, 602-603 (1987).
22. Laemmli, U. K. *Nature* **227**, 680-685 (1970).

Auxin induces rapid changes in phosphatidylinositol metabolites

C. Ettlinger & L. Lehle*

Fakultät für Biologie, Universität Regensburg, Universitätsstrasse 31, 8,400 Regensburg, FRG

The plant growth-hormone auxin (indole-3-acetic acid, IAA) is involved in regulating such diverse processes as cell elongation, cell division and differentiation. The sequence of events leading to the various phenomena is still poorly understood. Both changes in extra- and intracellular pH (refs 1–4) and selective transcription^{5,6} are known to be induced by auxin. Evidence for auxin receptors at the plasmalemma membrane has been reported⁷, but the signal transduction pathway is not known, for this nor for other plant hormones. In animal cells, hydrolysis of inositolphospholipids is a major mechanism for transmembrane signalling in response to external stimuli such as hormones, growth factors, neurotransmitters, antigens or light (reviewed in refs 8–11). Here we report that auxin can generate transient changes in inositol-1,4,5-trisphosphate (Ins(1,4,5)P₃) and inositol bisphosphate (InsP₂) within minutes in *Catharanthus roseus* cells arrested in G₁. These changes are accompanied by a redistribution within the polyphosphoinositide fraction. As the physiological response to auxin addition is to relieve the arrest in G₁, we suggest that these effects are an element in the signal transduction of this plant hormone.

The suspension culture from *C. roseus* used in this study requires auxin or the analogue 2,4-dichlorophenoxyacetic acid (2,4-D) as the only growth factor. Six days after subcultivation 95–98% of the cells arrest in G₁ (Fig. 1a). Resumption of cell division is strictly dependent on the presence of the hormone (Fig. 1b, c); during subsequent logarithmic growth about 50% of the nuclei are found to be in G₁. Arrested cells were metabolically labelled with [³H]myo-inositol in the absence of 2,4-D (see Fig. 2 legend). Of the radioactivity supplied, ~5% was found in the lipid fraction and 2% in the cell-wall polymer (total acid hydrolysis revealed that 95% of the cell-wall label was converted to arabinose and xylose). Less than 1% of the water-soluble fraction is composed of inositolphosphates, the rest being unmetabolized myo-inositol.

Figure 2a shows a typical thin-layer chromatography pattern of the polyphosphoinositides extracted from *C. roseus* membranes. The main products were identified as phosphatidylinositol, phosphatidylinositol phosphate and phosphatidylinositol bisphosphate (PtdIns, PtdInsP and PtdInsP₂) in the ratio of 92:5:2.5. A fourth dominant compound is assumed to be lyso-PtdIns on account of its migration behaviour. The radioactive phospholipids were characterized by co-chromatography with authentic standards (Amersham, Sigma). In addition PtdIns, PtdInsP and PtdInsP₂ were re-isolated from the plates, deacylated by transacylation to methylamine as described previously¹⁴ and the water-soluble radioactivity (recovery 95%) was analysed by high-pressure liquid chromatography (HPLC) as described in Fig. 2b. The deacylated compounds co-chromatographed with glycerol PIns (GroPIns), GroPIns4P, and GroPIns(4,5)P₂, respectively (data not shown). Reference compounds for column calibration were obtained by deacylating the commercially available radioactive lipids PtdIns, PtdIns4P and PtdIns(4,5)P₂ (Amersham). Total hydrolysis of the phospholipid fraction (2-N trifluoroacetic acid, 2 h at 120°C) gave only inositol as the radioactive compound.

Separation of water-soluble metabolites extracted from the cells was achieved by anion-exchange HPLC (Fig. 2b). Following published separation protocols^{15–17} and using authentic

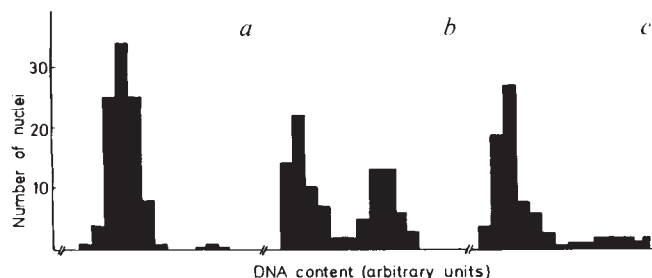


Fig. 1 Frequency distribution of nuclear DNA content from *Catharanthus roseus*. a, Cells cultured for 6 days in G₁. b, Cells recultured in the presence of 2,4-D for 24 h. c, Cells recultured without auxin but in the presence of the same amount of ethanol as used to dissolve the hormone. DNA content of Feulgen-stained nuclei was determined by scanning cytophotometry as described¹². Left peak, G₁ nuclei; right peak, G₂ nuclei. The cell line used was established by Dr J. Berlin (GBF, Braunschweig). Cells were grown in suspension in the dark on a gyratory shaker at 110 revs per minute (r.p.m.) at 29 °C in MS medium¹³ in the presence of 0.44 mg l⁻¹ 2,4-D and subcultivated every week.

Ins(1)P and Ins(1,4,5)P₃ (Amersham) and adenine nucleotide markers (AMP, ADP and ATP) for spiking, the radioactive inositol phosphates were identified as InsP, InsP₂ and Ins(1,4,5)P₃. The method used separates Ins(1,3,4)P₃ from Ins(1,4,5)P₃^{15–17}, but InsP and InsP₂ were not unambiguously identified as Ins(1)P and Ins(1,4)P₂.

Application of 2,4-D elicits a rapid, transient rise in Ins(1,4,5)P₃ and in InsP₂ (Fig. 3). The time-course of changes in polyphosphoinositides is characterized by a decline of radioactivity both in PtdInsP₂ and PtdInsP followed by a recovery and even a transient doubling. The percentage change in PtdIns is only small, but as it is the main labelled compound, this decrease nevertheless almost balances the change in radioactivity in the polyphosphoinositides. It should be emphasized, however, that it is not known whether equilibrium labelling of all pools has been reached and, thus, whether mass changes are measured. The biologically inactive auxin analogue 3,5-D did not trigger the hydrolysis of phosphoinositides, whereas indoleacetic acid caused the same changes as 2,4-D (data not shown). Li⁺ has been used to improve the sensitivity of the measurement of inositolphosphate formation, using its inhibitory effect on phosphatases^{18–21}. When 10 mM Li⁺ was added, a second trisphosphate isomer with the chromatographic behaviour of Ins(1,3,4)P₃ was observed (Fig. 2b insert; the compound co-elutes precisely with ATP). This compound is formed transiently after auxin stimulation with similar kinetics to that of Ins(1,4,5)P₂ (data not shown), perhaps indicating that the alternative pathway for metabolizing Ins(1,4,5)P₃ via formation of inositol 1,3,4,5-tetrakisphosphate found recently in animal tissues²² also exists in plants. Attempts to detect InsP₄ in the water-soluble material by HPLC, however, have so far failed.

Ins(1,4,5)P₃ is a second messenger releasing Ca²⁺ from internal stores in animal systems²³. Elevated levels of cytoplasmic Ca²⁺ in turn are thought to modulate a variety of developmental processes in many biological systems²⁴ including plants²⁵. It has been shown recently that Ins(1,4,5)P₃ causes a rapid release of Ca²⁺ from membrane vesicles from carrots and oat roots, respectively^{26–28}. We have found that a microsomal fraction derived from *C. roseus* cells responds similarly to the addition of Ins(1,4,5)P₃ and that Ins(1)P had no such effect (data not shown).

Our results support an auxin-triggered reaction sequence with an initial cleavage of PtdInsP₂ analogous to that in animal systems. The existence of enzymes assumed to be involved in the phosphatidylinositol cycle, like phospholipase C^{29,30}, pro-

* To whom correspondence should be addressed.

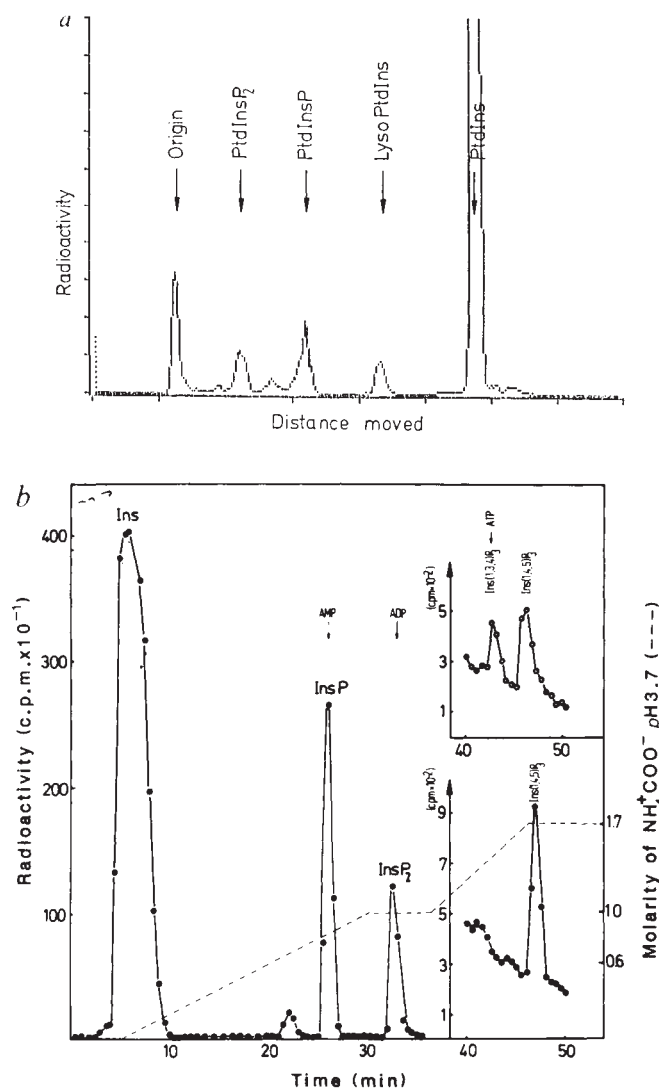
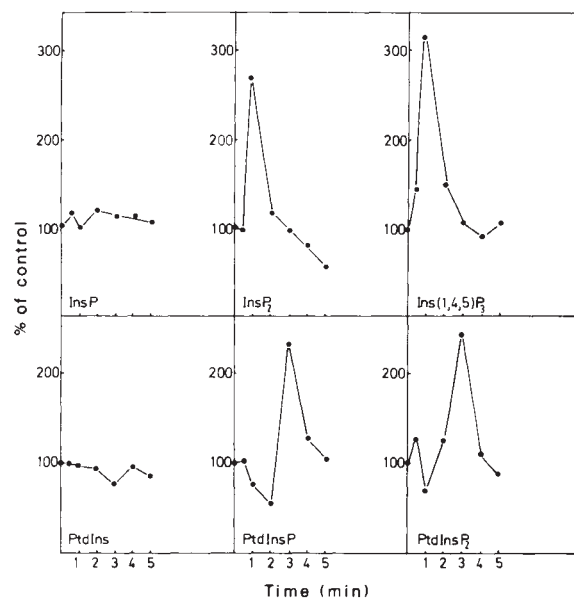


Fig. 3 Turnover of polyphosphoinositides in response to 2,4-D stimulation. A cell suspension was labelled as in Fig. 2 and divided into two samples. At time zero, one sample was stimulated by 2,4-D ($1 \mu\text{g ml}^{-1}$); the parallel control culture received ethanol in the same amount used to dissolve 2,4-D. Aliquots (3 ml) were removed at time zero, 30, 60, 120, 180, 240 and 300 s, and assayed for phosphoinositides and water-soluble metabolites as described in Fig. 2. Formation of products is expressed as percentage of the amount in control cells. Data from a single representative experiment are shown, out of three which gave essentially the same result. The 100% values correspond to the amounts given in the legend to Fig. 2.

tein kinase C^{31} and phosphoinositide kinases^{32,33} have been reported in plants. Also the occurrence of polyphosphoinositides and water-soluble inositol phosphates has been described³⁴⁻³⁶. So far, however, only an auxin-stimulated degradation of PtdIns to free inositol in isolated plant membranes has been demonstrated³⁷. Our data do not indicate that this is a primary response *in vivo*. Previously a correlation between the occurrence of polyphosphoinositides and the proliferation of *Catharanthus* cells has been reported³⁵. We suggest, therefore, that the observed fast breakdown of polyphosphoinositides (maximal

Fig. 2 (left) *a*, Analysis of [^3H]myo-inositol-labelled phospholipids by thin-layer chromatography; *b*, analysis of water-soluble inositol phosphates by HPLC.

Methods. *Catharanthus* cells in G_1 were collected by centrifugation and adjusted to a packed-cell volume of 40% using medium of a parallel culture of the same age. A 30-ml portion was then labelled with [^3H]myo-inositol ($10 \mu\text{Ci ml}^{-1}$, $17.9 \text{ Ci mmol}^{-1}$) for 3 h with continuous shaking, centrifuged, resuspended in fresh MS medium without 2,4-D but with 10 mM unlabelled myo-inositol and incubated for another 60 min (final volume as above). Samples (3 ml) were stopped by cooling down to 2°C within 30 s in precooled (-80°C) centrifuge tubes. Cells were collected by a 2 min centrifugation. Ice cold 10% trichloroacetic acid (TCA) was added and cells sonicated on ice for 2 min. Centrifugation gives a membrane/cell wall pellet and a soluble fraction. The pellet was washed with 5 ml of cold 5% TCA containing 1 mM Na-EDTA and finally with water. Phosphoinositide lipids were extracted with 5 ml chloroform:methanol:10N HCl (300:200:1.5) and subsequently with the same mixture in the ratio 400:200:1.5. To prevent auto-oxidation, 0.056% *t*-butyl-hydroxymethylphenol was included. The lipid extracts were combined and partitioned with water (to give chloroform:methanol:water 3:2:1), dried under nitrogen, redissolved in 0.2 ml chloroform:methanol 2:1 and the radioactivity determined in an aliquot by scintillation counting. TLC of lipids was performed on silica-gel-60 plates (Merck) in chloroform:methanol:13.5 N NH_3 :water in a ratio 45:30:3:5. Radioactive compounds were detected using a TLC analyser (Bertold). Spots corresponding to phosphoinositides were scraped off the plates and radioactivity measured by scintillation counting. Figure 2*a* shows an analysis of a 40- μl aliquot. The radioactivity (in d.p.m.) found on PtdIns was measured at 567,400; on PtdInsP, 28,830 and on PtdInsP₂, 15,915; the counting efficiency was 20%. Inositol phosphates were analysed after extracting TCA with diethylether by Partisil-SAX anion-exchange chromatography (Whatman). Ammonium formate (adjusted to pH 3.7 with phosphoric acid) was used for gradient elution with a flow rate of 1 ml min^{-1} . Adenine nucleotides were added as internal reference markers and their elution was detected at 254 nm. In *b*, the profile of an unstimulated extract is shown (total aliquot corresponding to 3 ml cell suspension). Eluted radioactivity was measured by liquid scintillation counting. After correcting for the gradient-dependent increase in quenching, the corrected values (d.p.m.) were InsP, 558,877; InsP₂, 387,875 and InsP₃, 3,381. To test the efficiency of extraction, labelled InsP₃ and PtdInsP₂ were added to unlabelled cells and taken through the procedure; recoveries were more than 90%. Pattern observed without LiCl (●) or in the presence of 10 mM LiCl (○, see insert) as described in the text.



response within minutes) may initiate a signal cascade and mobilize calcium, resulting in auxin-specific responses.

We thank Drs W Tanner and H. Beevers for stimulating discussions, advice and help in preparing the manuscript. 3,5-D was kindly supplied by Dr M Zenk. This work was supported by the Deutsche Forschungsgemeinschaft.

Received 29 July; accepted 20 November 1987.

- Hager, A., Menzel, H. & Krauss, A. *Planta* **100**, 47-75 (1971).
- Cleland, R. E. *A. Rev. Pl. Physiol.* **22**, 197-222 (1971).
- Felle, H., Brummer, B., Bertl, A. & Parish, R. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8992-8995 (1986).
- Taiz, L. *A. Rev. Pl. Physiol.* **35**, 585-657 (1984).
- Hagen, G., Kleinschmidt, A. & Guilfoyle, T. *Planta* **162**, 147-153 (1984).
- Theologis, A., Huynh, T. V. & Davies, R. W. *J. molec. Biol.* **183**, 53-68 (1985).
- Löbler, M. & Klämbt, D. *J. biol. Chem.* **260**, 9854-9859 (1985).
- Berridge, M. J. *Biol. Chem. Hoppe-Seyler* **367**, 447-456 (1986).
- Nishizuka, Y. *Nature* **308**, 693-698 (1984).
- Michell, B. *Nature* **319**, 176-177 (1986).
- Majerus, P. W. *et al. Science* **234**, 1519-1526 (1986).
- Eutlinger, C., Schindler, J. & Lehle, L. *Planta* **168**, 101-105 (1986).
- Murashige, T. & Skoog, F. *Physiologia Pl.* **15**, 437-497 (1962).
- Clarke, N. G. & Dawson, R. M. C. *Biochem. J.* **195**, 301-306 (1981).

- Irvine, R. F., Ånggörd, E. E., Letcher, A. J. & Downes, C. P. *Biochem. J.* **229**, 505-511 (1985).
- Turk, J., Wolf, B. A. & McDaniel, M. L. *Biochem. J.* **237**, 259-263 (1986).
- Hawkins, P. T., Stephens, L. & Downes, C. P. *Biochem. J.* **238**, 507-516 (1986).
- Berridge, M. J., Downes, C. P. & Hanley, M. R. *Biochem. J.* **206**, 587-595 (1982).
- Burgess, G. M., McKinney, J. S., Irvine, R. F. & Putney, J. W. Jr *Biochem. J.* **232**, 237-243 (1985).
- Bansal, V. S., Inhorn, R. C. & Majerus, P. W. *J. biol. Chem.* **262**, 9444-9447 (1987).
- Ackermann, K. E., Gish, B. G., Honchar, M. P. & Sherman, W. R. *Biochem. J.* **242**, 517-524 (1987).
- Irvine, R. F., Letcher, A. J., Heslop, J. P. & Berridge, M. J. *Nature* **320**, 631-634 (1986).
- Streb, H., Irvine, R. F., Berridge, M. J. & Schulz, I. *Nature* **306**, 67-69 (1983).
- Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315-321 (1984).
- Hepler, P. K. & Wayne, R. O. *A. Rev. Pl. Physiol.* **36**, 397-439 (1985).
- Drobak, B. K. & Ferguson, I. B. *Biochem. biophys. Res. Commun.* **130**, 1241-1246 (1985).
- Rincon, M. & Boss, W. F. *Pl. Physiol.* **83**, 395-398 (1987).
- Schumaker, K. S. & Sze, H. J. *J. biol. Chem.* **262**, 3944-3946 (1987).
- Irvine, R. F., Letcher, A. J. & Dawson, R. M. C. *Biochem. J.* **192**, 279-283 (1980).
- Helsper, J. P. F. G., DeGroot, P. F. M., Linskens, H. F. & Jackson, J. F. *Phytochemistry* **25**, 2053-2055 (1986).
- Elliott, D. & Skinner, J. D. *Phytochemistry* **25**, 39-44 (1986).
- Sandelius, A. S. & Sommarin, M. *FEBS Lett.* **201**, 282-286 (1986).
- Heim, S., Bauleke, A., Wylegalla, C. & Wagner, K. G. *Pl. Sci.* **49**, 159-165 (1987).
- Boss, W. F. & Massel, M. O. *Biochem. biophys. Res. Commun.* **132**, 1018-1023 (1985).
- Heim, S. & Wagner, K. G. *Biochem. biophys. Res. Commun.* **134**, 1175-1181 (1986).
- Morse, M. J., Crain, R. C. & Satter, R. L. *Pl. Physiol.* **83**, 640-644 (1987).
- Morre, D. J., Gripshover, B., Monroe, A. & Morre, J. T. *J. biol. Chem.* **259**, 15364-15368 (1984).

Functioning haemoglobin genes in non-nodulating plants

Didier Bogusz*†, Cyril A. Appleby*, Jörg Landsmann*‡, Elizabeth S. Dennis*, Michael J. Trinick* & W. James Peacock*

* Commonwealth Scientific and Industrial Research Organization, Division of Plant Industry, GPO Box 1600, Canberra, A.C.T. 2601, Australia

† ORSTOM Department MAA, 213 rue La Fayette, 75480 Paris, France

‡ Institute für Biochemie, Biologische Bundesanstalt, Messeweg 11/12, D-3300 Braunschweig, FRG

Haemoglobin has previously been recorded in plants only in the nitrogen-fixing nodules formed by symbiotic association between *Rhizobium* or *Frankia* and legume or non-legume hosts¹⁻⁴. Structural similarities amongst these and animal haemoglobins at the protein and gene level suggested a common evolutionary origin^{1,5-8}. This suggests that haemoglobin genes, inherited from an ancestor common to plants and animals, might be present in all plants. We report here the isolation of a haemoglobin gene from *Trema tomentosa*, a non-nodulating relative of *Parasponia* (Ulmaceae)⁹. The gene has three introns located at positions identical to those in the haemoglobin genes of nodulating plant species, strengthening the case for a common origin of all plant haemoglobin genes. The data argue strongly against horizontal haemoglobin gene transfer from animals to plants. The *Trema* gene has a tissue-specific pattern of transcription and translation, producing monomeric haemoglobin in *Trema* roots. We have also found that the *Parasponia* haemoglobin gene is transcribed in roots of non-nodulated plants. These results suggest that haemoglobin has a role in the respiratory metabolism of root cells of all plant species. We propose that its special role in nitrogen-fixing nodules has required adaptation of the haemoglobin-gene regulation pathway, to give high expression in the specialized environment of the nodule.

We have searched for haemoglobin genes in a number of plant species. Using a soybean haemoglobin complementary DNA probe, we were unable to detect (because of sequence divergence) the haemoglobin gene sequence even in *Parasponia*, a nodulating species known to produce haemoglobin⁶. We were able to isolate the *Parasponia* gene using a *Parasponia* cDNA clone identified with an oligonucleotide probe designed on the basis of amino-acid sequence data⁶. Because of the limitation of sequence divergence we first searched, using a *Parasponia* cDNA probe, for haemoglobin genes in non-nodulating plants in species of Ulmaceae, the family to which *Parasponia* belongs⁹.

We chose *Trema tomentosa*, taxonomically close to *Parasponia*, and *Celtis australis*, located in another section of the family. *Trema* species do not have nodulating roots in their native habitats, and we have shown that *T. tomentosa* is not capable of forming symbiotic associations with any of 98 representative strains of *Rhizobium* and *Bradyrhizobium*. We were unable to detect any nodulation in another *Trema* species, *T. orientalis*, or in *Celtis*⁹.

Southern hybridization showed haemoglobin-related sequences in both *Trema* and *Celtis*. The *Parasponia* haemoglobin cDNA segment hybridized to one band of *Trema* DNA following digestion with *EcoRI* or *BglII*. This suggests that there is a single haemoglobin-related segment in the genome of *Trema*, a situation comparable to *Parasponia* where there is only one haemoglobin gene. In *Celtis* the results indicate that there are either one or two genes present. Legume species have several haemoglobin genes and pseudogenes^{10,11}.

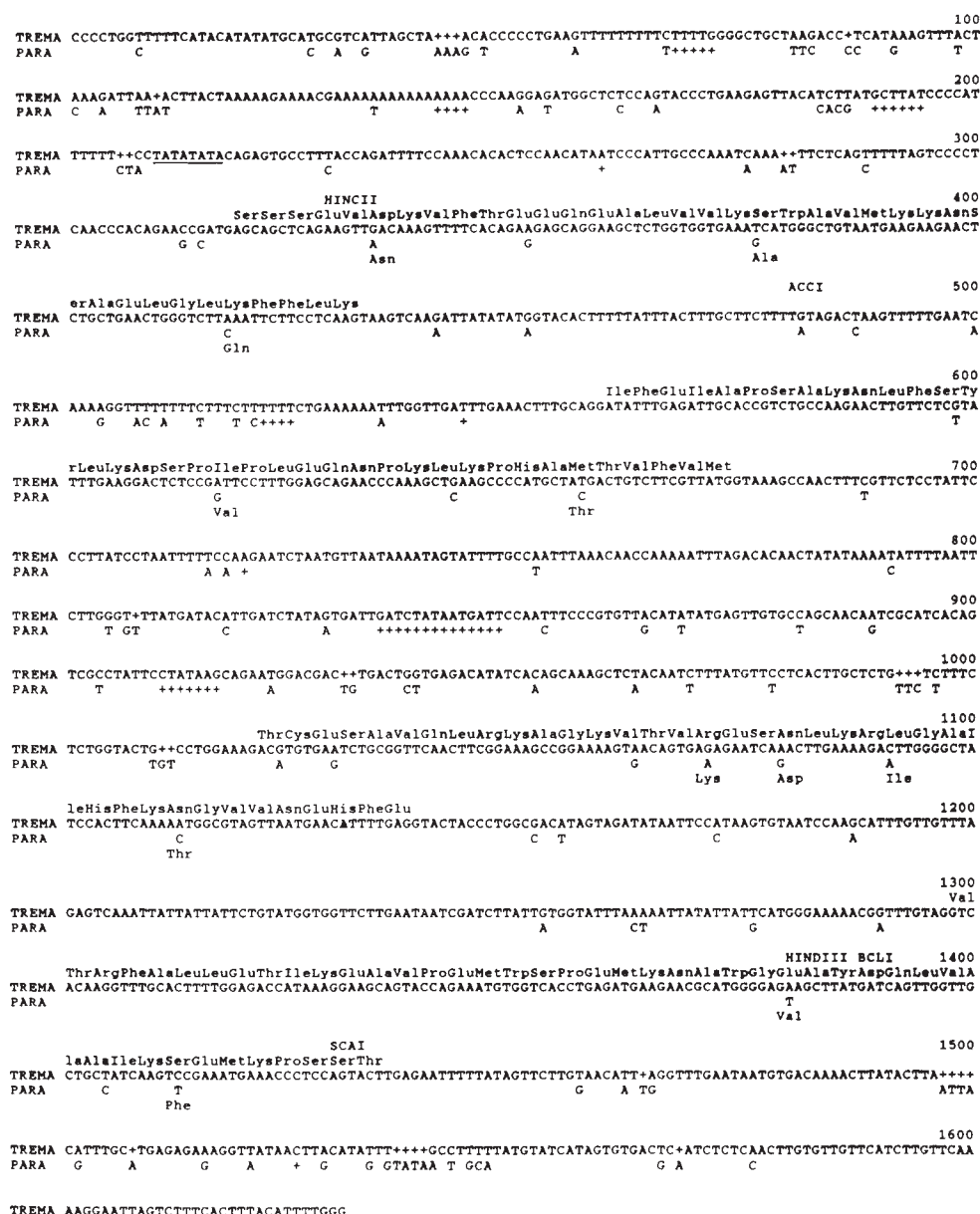
In *Trema*, we sought to determine whether the haemoglobin-related sequences correspond to a complete functional gene by isolating the appropriate DNA segment from the *T. tomentosa* genome (Fig. 1). Comparison of the *Trema* gene with the *Parasponia* gene⁶ identified four exons and three introns in identical positions. The *Trema* and *Parasponia* exon sequences have 93% nucleotide similarity and there is ~80% similarity in the introns, non-translated leader and 3'-untranslated sequences of the genes.

A haem-polypeptide with amino-acid sequence deduced from the *Trema* gene sequence (Fig. 1) would have the properties expected of a functional haemoglobin. It has a chain length (161 residues) identical to *Parasponia* haemoglobin, and has the 30 amino-acid residues common to all plant haemoglobins whose structures are known^{6-8,12,13}. These include the residues equivalent to proximal and distal His, Pro C2 and Phe CD1, which are invariant or highly conserved in all functional haemoglobins. Of the 11 amino acids not identical between the *Parasponia* and *Trema* polypeptides (Fig. 1), seven are not found in any other plant haemoglobin. These are Asp 6, Lys 34, Ile 58, Met 71, Arg 93, Asn 96 and Ser 154, with only Arg 93 having the potential to cause instability in *Trema* haemoglobin; compare with refs 8 and 14. We conclude that the *Trema* gene is likely to be functional, and that its product could participate in oxygen transport¹.

We analysed *Trema* tissue for RNA transcripts and for protein products of the gene (Fig. 2). Poly(A)⁺ RNA extracted from *Trema* roots (lane 4) hybridized to *Parasponia* haemoglobin cDNA but no haemoglobin messenger RNA was detected in *Trema* leaf extracts (lane 5). The single band in root RNA indicates an mRNA of 0.7 kilobases (kb), identical in length to the haemoglobin mRNA in *Parasponia* nodules (lane 1).

Fig. 1 Alignment of the *T. tomentosa* (TREMA) and *P. andersonii* (PARA) haemoglobin genes. The amino-acid translation is given above and below the coding sequence. That of *Trema* is given in full. Only variant nucleotides and amino acids are given for *Parasponia*. Non-coding sequences are aligned to give maximum base-for-base similarity. The proposed TATA box is underlined (____). Crosses (+++) represent the sequences deleted or inserted.

Methods. A genomic library in λ EMBL4 was made from a 4–7 kb *Bgl*III-fraction of *Trema* total DNA. About 2×10^5 recombinant phages were screened with 32 P-nick-translated *Parasponia* haemoglobin cDNA probe pL209 (ref. 6). Six genomic clones were isolated and one of them was subcloned successively in pUC18 and pBGS18 (ref. 19). The sequence was determined by subcloning fragments into M13 (mp18,19) phages and sequencing by the dideoxy-chain termination method²⁰.



We also detected haemoglobin mRNA in non-nodulated *Parasponia* roots (Fig. 2, lane 2), the level of mRNA being about 1,000-fold less than that found in *Parasponia* nodule tissue (lane 1). No signal was found in *Parasponia* leaves (lane 3). The *Parasponia* haemoglobin gene thus has two differently-regulated patterns of expression, a high level of transcription in nodules and a low level of transcription in normal root tissue. We did not detect any haemoglobin mRNA in non-nodulated soybean roots with either soybean or *Parasponia* haemoglobin cDNA probes; compare with refs 15 and 16. If there is expression of haemoglobin genes in non-nodule tissue of soybean, either it is at a level too low to be detected or it is from a non-symbiotic gene sufficiently divergent in sequence that no cross hybridization occurs with the probes we used.

We fractionated *Trema* root extracts to look for haemoglobin protein. Western blot analysis (Fig. 3) showed that the haemoglobin mRNA is translated in *Trema* roots and that the native product is a monomer as it is in nodules of legumes and *Casuarina*. In *Parasponia* nodules the native haemoglobin is a dimer^{2,17} and this is also true for the haemoglobin found in normal (non-nodulated) root tissue. Analysis of the root haemoglobin of *Parasponia* will determine if the monomer form is root-specific for that genus.

Our data suggest that all plants have haemoglobin genes, and

Fig. 2 Northern blot hybridization of *Parasponia* haemoglobin cDNA (pL209) (ref. 6) to total RNA isolated from *Parasponia* nodules (lane 1) and to poly(A)⁺ RNA isolated from *Parasponia* roots (lane 2), *Parasponia* leaves (lane 3), *Trema* roots (lane 4) and *Trema* leaves (lane 5).

Methods. Uninoculated roots of *Parasponia andersonii* were obtained from plants grown aseptically on nutrient agar. Total RNA was extracted and poly(A)⁺ purified according to standard procedures²¹. About 2 μ g of total *Parasponia* nodule RNA and 5 μ g of poly(A)⁺ RNA was subjected to electrophoresis on a 1% formaldehyde agarose gel²¹, blotted and hybridized. The probe was labelled with 32 P by an oligolabelling procedure²². The blots were hybridized at 42 °C as described⁶ but with 10% dextran sulphate added. The filters were washed twice in a 2 $\times$ SSC, 0.1% sodium dodecyl sulphate (SDS) at room temperature for 15 min, then once in 0.1 $\times$ SSC, 0.1% SDS at room temperature for 30 min before autoradiography.

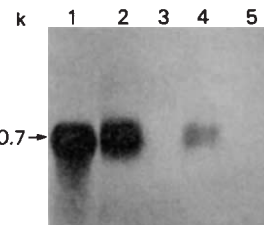


Fig. 3 Western blot identification of a native monomeric haemoglobin from *T. tomentosa* root tips, following non-denaturing molecular-exclusion chromatography on Sephacryl S200. Lane 1, pure *Parasponia* haemoglobin subunit, 0.1 µg; lane 2, the 'haemoglobin dimer' fractions $M_r \sim 39K$, ex S200; lane 3, the 'peroxidase' $M_r \sim 33K$ fraction ex S200; lane 4, the 'haemoglobin monomer' fraction, $M_r \sim 18K$, ex S200; lane 5, the 'cytochrome c' $M_r \sim 13K$ fraction ex S200. The arrow indicates subunit $M_r \sim 19K$. A significant interaction with the *Parasponia* haemoglobin antiserum occurred only in the 'haemoglobin monomer' fraction (lane 4). We consider that the faint bands seen in the other lanes result from adsorption of *Trema* monomeric haemoglobin to other proteins during the non-denaturing chromatography.



Methods. Frozen tissue (18 g) was extracted anaerobically and chromatographed on a calibrated column of Sephacryl S200 (Pharmacia, post 1980 batch) by a non-denaturing procedure^{2,4}. The S200 effluent fractions expected to contain native dimeric haemoglobin ($M_r \sim 39K$)^{2,17}, peroxidase ($M_r \sim 33K$)¹², native monomeric haemoglobin ($M_r \sim 18K$)⁴ and cytochrome c ($M_r \sim 13K$)¹² were separately pooled and concentrated to 1 ml. Residual polyvinyl pyrrolidone⁴ was precipitated at 0.45 ammonium sulphate saturation, then putative haemoglobin and other proteins precipitated at 0.8 ammonium sulphate saturation. The redissolved protein pellets were water-washed and reconcentrated then proteins reprecipitated with 80% v/v acetone at $-80^\circ C$, and redissolved to $\sim 100 \mu l$ in Western blot digestion buffer⁴. Samples were applied to SDS-polyacrylamide gel lanes before Western blot analysis⁴ using IgG from rabbit anti-*Parasponia* haemoglobin serum, followed by protein A-peroxidase staining.

imply that haemoglobin has a function, presumably associated with oxygen transport, in cells of normal roots. In *Parasponia*, the product of a single gene sequence appears to have both root and nodule functions, but the level of gene expression is low in roots compared with nodules. We have not been able to detect haemoglobin mRNA in soybean roots but the possibility remains that there is a separate haemoglobin gene for the root function in legumes. Clearly if haemoglobin genes are present in all plants and encode proteins which function in root oxygen transport, then it is unnecessary to invoke horizontal gene transmission^{5,18} to explain the sparsely-scattered appearance of symbiotic haemoglobins in nitrogen-fixing nodules¹⁻⁴. Instead, the ubiquitous root haemoglobin gene may be the progenitor of the nodule genes. In this case we might expect that the nodule gene has evolved differently in each group of nodulating plants. In some species, for example *Parasponia*, the nodule and root haemoglobin functions may be provided by alternative regulation of the one gene, whereas in legumes, gene duplication may have been involved in the evolution of nodule-specific genes.

We thank Dr Don Spencer for help with Western blot analysis, Drs Jeff Ellis and Danny Llewellyn for helpful suggestions, Paul Hadobas for plants and Georgina Koci, Janice Norman and Judy Flanagan for technical assistance. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00296.

Note added in proof: Western blot analyses of *Parasponia* root haemoglobin showed it also to be a dimer.

Received 10 September; accepted 23 November 1987.

1. Appleby, C. A. *Rev. Pl. Physiol.* **35**, 443-478 (1984).
2. Appleby, C. A., Tjepkema, J. D. & Trinick, M. J. *Science* **220**, 951-953 (1983).
3. Tjepkema, J. D. *Can. J. Bot.* **61**, 2924-2929 (1983).
4. Fleming, A. I., Wittenberg, J. B., Wittenberg, B. A., Dudman, W. F. & Appleby, C. A. *Biochem. biophys. Acta* **911**, 209-220 (1987).
5. Hyldig-Nielsen, J. J. *et al. Nucleic Acids Res.* **10**, 689-701 (1982).
6. Landsmann, J. *et al. Nature* **324**, 166-168 (1986).

7. Fleming, A. I. *et al. in Proc. 8th Australian Nitrogen Fixation Conf.* (eds Wallace, W. & Smith, S. E.) 85-86 (Australian Institute of Agricultural Science, Parkville, 1986).
8. Dikerson, R. E. & Geis, J. *Hemoglobin: Structure, Function, Evolution and Pathology* (Benjamin/Cummings, California, 1983).
9. Akkermans, A. D. L., Abdulkadir, S. & Trinick, M. J. *Plant and Soil* **49**, 711-715 (1978).
10. Bojsen, K., Abildsten, D., Jensen, E. Ø., Paludan, K. & Marcker, K. A. *EMBO J.* **2**, 1165-1168 (1983).
11. Lee, J. S., Brown, G. G. & Verma, D. P. S. *Nucleic Acids Res.* **11**, 5541-5553 (1983).
12. *National Biological Research Foundation PIR Protein Sequence Database* release 12 (17 March, 1987).
13. Kortt, A. A., Strike, P. M., Bogusz, D. & Appleby, C. A. *Biochem. Int.* **15**, 509-516 (1987).
14. Fermi, G. & Perutz, M. F. *Haemoglobin and Myoglobin* (Clarendon, Oxford, 1981).
15. Marcker, A., Lund, M., Jensen, E. Ø. & Marcker, K. A. *EMBO J.* **3**, 1691-1695 (1984).
16. Jensen, J. S., Marcker, K. A., Otten, L. & Schell, J. *Nature* **321**, 669-674 (1986).
17. Wittenberg, J. B., Wittenberg, B. A., Gibson, Q. H., Trinick, M. J. & Appleby, C. A. *J. biol. Chem.* **261**, 13624-13631 (1986).
18. Runnegar, B. N. *J. molec. Evol.* **21**, 33-41 (1984).
19. Spratt, B. G., Hedge, P. J., Heesen, S. T., Edelman, A. & Broome-Smith, J. K. *Gene* **41**, 337-342 (1986).
20. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
21. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
22. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).

The mast cell binding site on human immunoglobulin E

Birgit Helm*, Philip Marsh*, Donata Vercelli†, Eduardo Padlan‡, Hannah Gould* & Raif Geha†

* Department of Biophysics, King's College, 26-29 Drury Lane, London WC2B 5RL, UK

† Division of Allergy and Immunology, Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

Antibodies of the immunoglobulin E isotype sensitize mast cells and basophils for antigen-induced mediator release by binding through the Fc portion to a high-affinity receptor ($Fc_\epsilon R1$, $K_a = 10^9 M^{-1}$) on the cell surface^{1,2} causing the clinical manifestations of type I hypersensitivity. As the amino acid sequence of the human epsilon chain is now known³, attempts have been made to map the $Fc_\epsilon R1$ binding site on IgE to a fragment smaller than Fc_ϵ using proteolytic cleavage products, none of which proved to be active³. Cleavage between the $C_\epsilon 2$ and $C_\epsilon 3$ domains released two inactive fragments, suggesting that the junction between these segments could be important in receptor binding³. This region is protected against protease digestion in the rat IgE complex with the receptor of rat basophilic leukaemia cells⁴. Here we report the mapping of the mast cell receptor binding site on human IgE to a sequence of 76 amino acids at the $C_\epsilon 2/C_\epsilon 3$ junction. Recombinant peptides containing this sequence inhibit passive sensitization of skin mast cells *in vivo* and sensitize mast cells to degranulation by anti-IgE *in vitro* almost as efficiently as a myeloma IgE. Fragments containing the separate domains are inactive. Additional sequences are required for rapid assembly of fragments into disulphide-linked dimers, suggesting that a single chain can form the active site. In a three-dimensional model of the human Fc_ϵ , the two identical segments are far apart. Each folds to generate a cleft between the $C_\epsilon 2$ and $C_\epsilon 3$ domains on the surface of the Fc_ϵ . The docking of IgE on to mast cells could take place within this cleft.

The bacterial expression product of the whole recombinant Fc_ϵ (rFc_ϵ) has been shown to be as active as a myeloma IgE in binding assays *in vivo* and *in vitro* and in assays of biological activity⁵⁻⁹. Here we extended this approach to the expression of smaller fragments of the ϵ -chain gene and assay their capacity to inhibit sensitization of mast cells by IgE antibodies *in vivo* and to sensitize basophils to degranulation by anti-IgE *in vitro*. The recombinant Fc_ϵ forms a disulphide-linked dimer on oxidation *in vitro*, resembling native Fc_ϵ (refs 5, 7), so we have examined the dimerization of the smaller peptides in relation to their activity. This information has been applied with a model of the three-dimensional structure of the human Fc_ϵ (ref. 10)

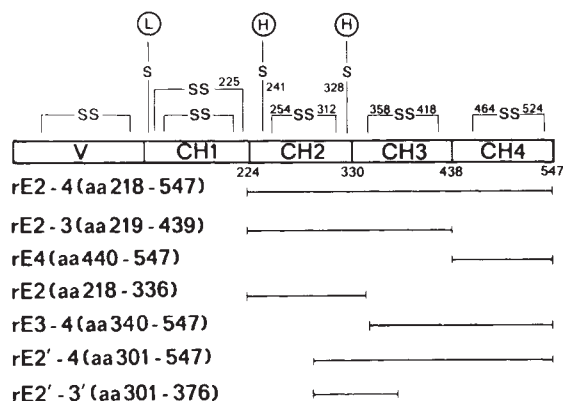


Fig. 1 Genetic expression products used in mapping the mast cell receptor site on the human epsilon chain. Above, covalent structure of the human ϵ -chain, indicating the positions of the intra-chain disulphide bonds (S-S) and inter-heavy (S-H) and heavy-light (S-L) chain disulphide bonds and the boundaries of the five structural domains, corresponding to exons in the variable (V) and constant (C) regions of genomic DNA (VH, CH1-4), from the N-terminus (left) to the C-terminus (right), drawn to scale. Amino acids are numbered according to ref. 3. Below, Peptides synthesized by *Escherichia coli* expression of gene fragments encoding the indicated protein sequences. The recombinant epsilon (rE) peptide names correspond roughly to the domains they contain, with primed numbers indicating a truncated domain. Gene constructions will be described elsewhere (P. Marsh, in preparation).

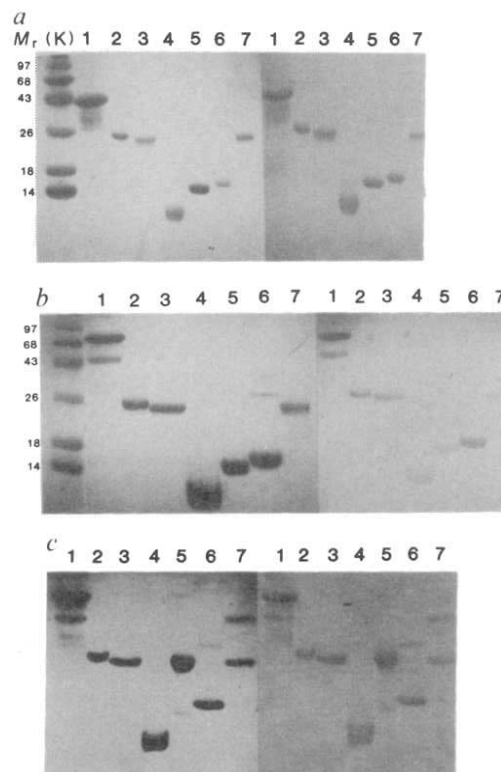


Fig. 2 Polyacrylamide gel electrophoresis of recombinant epsilon chain peptides. Peptides were isolated from *E. coli*, purified by affinity chromatography and subjected to a procedure used for refolding of immunoglobulin chains, see Methods. The resulting peptides were analysed by polyacrylamide gel electrophoresis under three sets of conditions: *a*, in SDS under reducing conditions; *b*, in SDS under non-reducing conditions; *c*, in non-denaturing gels. Duplicate gels were either stained with Coomassie blue (left) or transferred to nitrocellulose and assayed with a polyclonal antibody against human IgE (right). The peptides were loaded in the same order on each gel: rE2-4 (lane 1), rE2'-4 (lane 2), rE2-3 (lane 3), rE2'-3' (lane 4), rE4 (lane 5), rE2 (lane 6), rE3-4 (lane 7). **Methods.** The peptides were isolated from *E. coli* and affinity-purified with an appropriate antibody covalently linked to Sepharose 4B as described previously⁶. In the case of rE4, a rabbit antiserum directed against a synthetic oligopeptide with the sequence of the ϵ -chain from Leu 509-Leu 524 inclusive was used. After affinity purification, the peptides were unfolded, reduced, refolded and oxidized²³. Electrophoresis²⁴ was performed within 24 h of preparation; aliquots frozen immediately after preparation and analysed within the same interval gave identical results. Reduced samples were treated with 1% β -mercaptoethanol, 8 M urea and 1% SDS and boiled for 3 min before SDS-PAGE and non-reduced samples were prepared similarly but without β -mercaptoethanol treatment. Gels were stained with Coomassie blue R-250 and a duplicate gel analysed by Western blotting. After transfer to nitrocellulose, the epsilon peptides were reacted with a rabbit antiserum to purified human IgE (Dakopatts) and the rabbit antibody detected with a peroxidase-labelled swine anti-rabbit IgG antiserum (Dakopatts).

to predict the properties of the binding site.

The recombinant peptides are shown in Fig. 1, and their purity and monomer molecular weights are established by SDS-polyacrylamide gel electrophoresis (PAGE) (Fig. 2*a*). The molecular weights corresponded to those defined by the cloned DNA sequences. SDS-PAGE under non-reducing conditions (Fig. 2*b*) showed that >90% of the rFc_ε (rE2-4 in the present study) formed dimers, and that under the same conditions none of the shorter peptides (rE2, rE2-3, rE2'-4 or rE2'-3') formed inter-chain disulphide bonds, although after one week at 4 °C there was ~10% self-association for rE2 and 40% for rE2'-4 and rE2'-3', and rE2-3 remained as monomer (results not shown). Figure 2 shows that, despite the absence of interchain disulphide bonds, several of the peptides could form non-covalent dimers in non-denaturing conditions: the relative mobilities indicate that rE2-4 and rE4 are mainly dimers, rE3-4 is a 1:1 monomer-dimer mixture, and the remainder (rE2'-4, rE2'-3' and rE2-3) are predominantly monomers. Comparable molecular weights were derived from plots of relative mobility against acrylamide concentration¹¹ (results not shown).

The inhibition of passive cutaneous anaphylaxis in human skin, known as the Prausnitz-Kustner (P-K) reaction¹², was used to assay the biological activity of the recombinant peptides. A single serum (EC) which contained IgE antibodies to ragweed was used to sensitize skin sites of a single normal subject, and the skin sites challenged with ragweed antigen 48 hours later⁸. In preliminary experiments, pre-injection one hour before the serum of myeloma IgE(PS) or of rE2-4, rE2'-4, rE2-3 or rE2'-3' at 1 μ g ml⁻¹, completely blocked the P-K reaction induced by a 1:100 dilution of serum EC containing 5×10^{-11} M IgE. In diluent-treated sites, the area of wheal and erythema were respectively 87 ± 27 mm² and 416 ± 63 mm² ($n=6$). In contrast, fragments rE4 ($n=5$), rE2 ($n=6$) and rE3-4 ($n=4$) did not significantly diminish the areas of the wheal (75 ± 31 , 99 ± 22 and 80 ± 15 mm² respectively) and erythema (428 ± 56 , 387 ± 43 and 392 ± 74 mm² respectively), nor did an equimolar mixture of rE2 and rE3-4. The inhibition was specific, for when skin sites were injected with serum E.C. one hour before injection of the fragments, no inhibition was observed.

The potency of the active peptides in inhibiting the P-K reaction was quantified by assaying a logarithmic dilution series

of each (Table 1). The results for the smallest peptide rE2'-3', relative to a myeloma IgE(PS), are shown in Fig. 3*a*. The peptide is ~30% as effective as IgE on a molar basis, assuming the peptides are active as monomers. Table 1 shows that all the peptides containing the sequence Gln 301-Arg 376 were comparable in their inhibition of the P-K reaction. We measured the duration of the effect of the smallest fragment rE2'-3' after injection into normal skin. Skin sites remained refractory to passive sensitization for 9 and 14 days after injection with 0.1 and 1 μ M rE2'-3' respectively, comparing favourably with IgE(PS) protection of skin mast cells for 14 and 19 days at the same concentrations⁸.

The peptide rE2'-3' also retains the target-cell-triggering

activity characteristic of cross-linked native IgE. Peripheral blood basophils, when sensitized with rE2'-3' or IgE(PS), release 28% and 30% of total basophil protease activity respectively upon challenge with affinity-purified rabbit antibody raised against rE2'-3' (Fig. 3b) at the same optimal rabbit antibody concentration for both rE2'-3' and myeloma IgE. The efficacy of antibodies in triggering cell degranulation in this assay signifies that a part of the sequence in Gln 301-Arg 376 is accessible to antibody on the surface of the IgE receptor complex. These observations do not substantiate earlier conclusions that a decapeptide in C_ε4 is important for IgE-mediated target-cell triggering¹³.

Thus only the Gln 301-Arg 376 sequence is sufficient for high-affinity binding to the receptor on mast cells and basophils. This sequence should probably reside in a single polypeptide chain to be active: Gln 301-Val 336 (contained in rE2) and Leu 340-Arg 376 (contained in rE3-4) are inactive. None of the sub-fragments of the human Fc_ε studied previously had this complete sequence, although cleavage with pepsin at Ser 338 destroyed the receptor binding site, consistent with our observations. We also conclude that ε-chain dimers are not necessary for receptor binding in that rE2-3, which remained a monomer in all conditions tested, was no less active than other peptides that dimerized efficiently.

Synthetic peptides corresponding to surface-accessible sequences within the C_ε3 and C_ε4 domains of rat IgE compete with rat ¹²⁵I-labelled IgE for binding to rat mast cells¹⁴, the most active peptide being 1,000-fold less potent than unlabelled rat IgE. A human pentapeptide inhibitor of the allergic reaction requires >10⁵ the concentration of IgE for 50% inhibition, and its effect is nonspecific¹⁵. This contrasts with our recombinant peptides with at least 30% of IgE activity *in vitro* and *in vivo*.

C_ε4 is not essential for receptor binding or for triggering cell degranulation, but is evidently important in ε-chain assembly into the dimer found in IgE. Two of the four recombinant peptides containing this domain, rE2-4 and rE4, dimerized readily, but the peptides lacking C_ε4 did not. The peptides containing the C_ε2 domain either alone (rE2) or with C_ε3 (rE2-3), did not dimerize under the standard conditions. Furthermore, rE2-4 formed a covalent dimer, whereas rE2'-4 did not, despite the presence of both C_ε4 and one of the two cysteines (Cys 328, but not Cys 241) involved in interchain disulphide bond formation in C_ε2. Thus C_ε4 nucleates the ε-chain dimer formation in IgE, facilitating interchain disulphide bonding in C_ε2 to lock the chains together. The homologous C_ε3 domain functions similarly in the assembly of IgG (ref. 16).

Two similar models of the three-dimensional structure of the human Fc_ε have been proposed based on the sequence homology between the individual domains and those in IgG1 (refs 10, 17), for which a high-resolution crystal structure is available¹⁸. Our model shows the domains in different colours and the Gln 301-Arg 376 sequence highlighted on one of the ε-chains in Fig. 4. The peptide is predicted to contain the last two β-strands of C_ε2 and the first three of C_ε3. The two strands in C_ε2 are on the side exposed to solvent, the first two strands in C_ε3 face away from solvent and, by homology with IgG (ref. 18), would be extensively masked by carbohydrate; the third strand in C_ε3 faces the solvent. The carbohydrate present on Asn 371 (Fig. 4) in native IgE (ref. 20) should render the C-terminal part of the sequence inaccessible. Human IgE also has carbohydrate on Asn 265 (ref. 20). This residue is outside the sequence of interest here, but the proximity of Asn 265 to the C_ε2 domain of the peptide (Fig. 4) would mask the bend in the sequence, restricting the putative site of interaction with receptor to the regions on either side of this bend. The portions of the sequence that are not buried in the interior of the IgE form a cleft between the C_ε2 and C_ε3 domains: one side of the cleft is defined by two segments of the sequence in C_ε2 and the other by one in C_ε3. The α-chain of the IgE receptor molecule protrudes from the mast cell membrane¹⁹, and probably binds

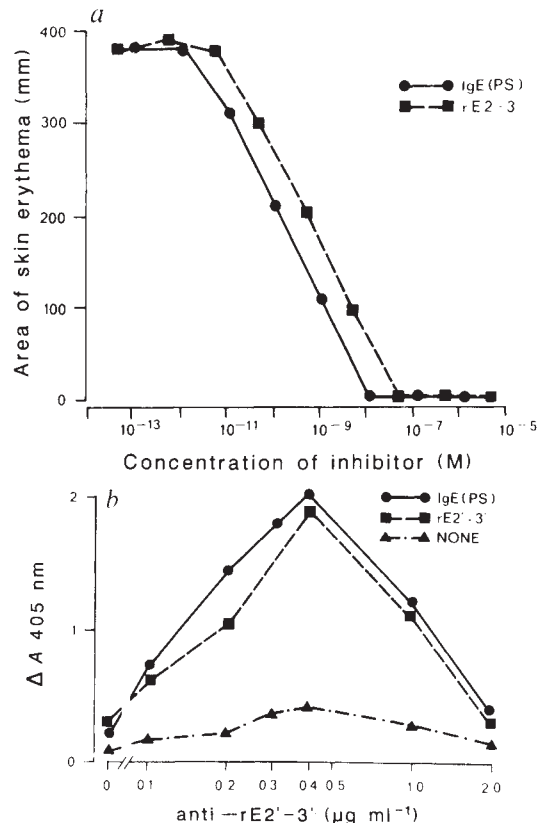


Fig. 3 Biological activity of rE2'-3'. The activity of rE2'-3' was compared with that of IgE(PS) *a*, blocking the P-K reaction *in vivo* and *b*, in a protease release assay (for sensitization of basophils) *in vitro*.

Methods. *a*, The methods used for the P-K test are described in the legend to Table 1. Here rE2'-3' and IgE(PS) are compared as a function of molar concentration, assuming relative molecular masses of 9,000 (monomeric) and 180,000 respectively. *b*, Leukocytes isolated from the plasma of a non-atopic individual (P.M.) were concentrated to 0.5–2.0 × 10⁷ cells (0.4–1.6 × 10⁵ basophils) per ml by dextran sedimentation. Experimental details accompany the Photometric Allergy Degranulation test kit (Boehringer-Mannheim). For passive sensitization, 0.5 ml aliquots were incubated at 4 °C for 90 min with 2 μg ml⁻¹ IgE(ND) or 0.8 μg ml⁻¹ rE2'-3'. The cells were then washed with RPMI 1640 medium resuspended in 10 mM HEPES-buffered Tyrode solution, and incubated for 30 min at 37 °C with increasing concentrations of affinity-purified rabbit anti-rE2'-3', obtained by passing a rabbit anti-human epsilon antibody over an affinity column of rE2'-3'. The reaction was stopped by addition of ice-cold phosphate-buffered EDTA and centrifuged immediately for 5 min at 900g. The spectrophotometric change in absorbance at 405 nm (A₄₀₅) of the supernatant caused by the release of the basophil cell-specific protease, liberating p-NO₂-anilide from chromozym TH, was measured. Curves were obtained for control cells, cells incubated with rE2'-3', and cells incubated with IgE(PS), as indicated. Each point represents the mean of three determinations. Total protease activity present in aliquots of lysed cells resulted in an increase in A₄₀₅ of 6.7 units in 30 min.

within this cleft. The two peptides rE2 and rE3-4 together contain all of the sequences exposed in the cleft, but are inactive in binding, as would be expected should high-affinity binding require contacts on both sides of the cleft. We cannot exclude the possibility that the three amino acids separating the peptides tested (amino acids 337–339) are critical for binding, but consider this unlikely as they face away from solvent in the proposed structure (Fig. 4). The interaction of protein A with IgG also requires contiguous elements from two immunoglobulin domains^{18,21}.

The Fc_ε contains two identical ε-chain chains lying on a twofold dyad axis of symmetry. The two Gln 301-Arg 376

Fig. 4 Three-dimensional model of the mast cell receptor-binding site on human IgE (stereo-pair). Proposed model of human IgE Fc (ref. 10) showing the putative location of peptide Gln 301–Arg 376 (white segment in the ϵ -chain on the left (blue)). The C₂ (red), C₃ (green) and C₄ (magenta) domains are differentiated in the ϵ -chain on the right. The carbohydrate substituents between the two C₃ domains are shown in yellow. The asparagine residues, Asn 265 and Asn 371, bearing carbohydrate in human IgE Fc are shown in yellow and are labelled in the ϵ -chain on the right. The beginning and end of the Gln 301–Arg 376 segment and two of the three residues, 337 and 338, that separate the rE2 and rE3-4 peptides, are labelled in the ϵ -chain on the left.

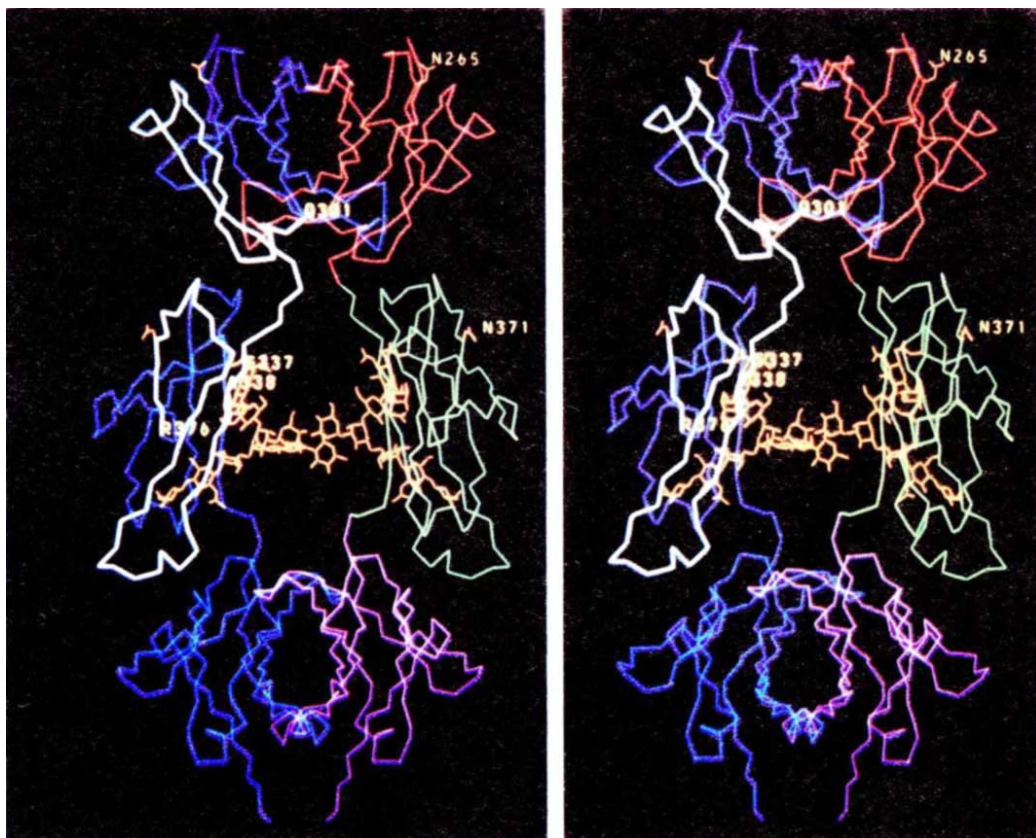


Table 1 Relative activity of recombinant IgE peptides in the inhibition of the P-K reaction

Source	First experiment (5×10^{-11} M IgE)		Second experiment (2.5×10^{-10} M IgE)	
	Concentration*	Potency	Concentration*	Potency
IgE(PS)	2×10^{-10}	100	2×10^{-9}	100
rE2-4	4×10^{-10}	50	4×10^{-9}	50
rE2'-4	5×10^{-10}	40	4×10^{-9}	50
rE2-3	5×10^{-10}	40	6×10^{-9}	33
rE2'-3'	6×10^{-10}	33	5×10^{-9}	40

A single subject (R.G.) was used for passive sensitization. The serum IgE of this subject was 4 IU ml^{-1} ($\sim 10 \text{ ng ml}^{-1}$). The sensitizing serum (E.C.) contained 380 IU ml^{-1} of IgE (912 ng ml^{-1}), of which 8.7% was directed against ragweed antigen, as determined by the specific drop in serum IgE following absorption of the serum over a Sepharose 4B ragweed antigen column, compared with a control Sepharose 4B human serum albumin. Serum E.C. was free of detectable hepatitis B antigen and of antibodies to that antigen and to human immunodeficiency virus (HIV). Serum E.C. was obtained in 1983 and the donor is currently healthy and HIV-antibody negative. Epsilon chain fragments were injected into skin sites 1 h before the injection of serum E.C. Skin sites were challenged 48 h later with ragweed antigen ($1,000$ protein nitrogen units ml^{-1} of a mixture of giant and short ragweed; Hollister-Stier, Spokane, Washington). After 20 min the skin sites were examined for wheal and erythema. The surface area of the reactions was estimated as follows: transparent tape was used to transfer the outline for the reactions to paper, which was cut and weighed. The area was read from a standard curve. All injections (0.02 ml) were intradermal. In each experiment a set of skin sites was also sensitized with diluent. None of these sites showed a wheal or flare reaction when challenged with ragweed antigen. Diluent consisted of 0.15 M NaCl and 0.03% human serum albumin. The molarities of the ϵ -chain peptides were calculated taking into account the proportion of dimers and monomers in each preparation. Monomers were included in the calculations because rE2-3, which never dimerized detectably, was a potent inhibitor of the P-K reaction (see above). Each fragment was used over a range of 10^{-13} M– 10^{-6} M in tenfold increments.

* Molarity required for 50% of the P-K reaction.

sequences make no close contacts in this structure. At the point of closest approach, at the α -carbons of Arg 334, the ϵ -chains are 1.1 nm apart. This topography probably precludes a mechanism of receptor binding involving equivalent contacts of the receptor on the two ϵ -chains; this would entail entry from opposite sides of the IgE molecule, but the high-affinity receptor on rat basophilic leukaemia cells contains only a single binding subunit²². Occupancy of one of two equivalent sites could be sufficient for the effector function of IgE.

We thank Drs J. Hakimi and W. Danhoff (Hoffmann LaRoche) for the antibody used in purifying rE4, and Drs H. Metzger, D. R. Davies and T. E. Creighton for valuable discussions. The work was supported by grants from the USPHS, the National Foundation, March of Dimes, the MRC, the CRC and the Research Corporation Ltd.

Received 4 August; accepted 16 November 1987.

1. Stanworth, D. R., Humphrey, J. H., Bennich, H., Johansson, S. G. O. *Lancet* **2**, 17–18 (1968).
2. Ishizaka, K., Ishizaka, T. & Lee, E. H. *Immunochimistry* **7**, 687–702 (1970).
3. Dorrington, K. J. & Bennich, H. *Immun. Rev.* **41**, 3–25 (1978).
4. Perez-Montfort, R. & Metzger, H. *Molec. Immun.* **19**, 1113–1125 (1982).
5. Kente, J. H., Helm, B. A., Ishizaka, T., Cattini, P. & Gould, H. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2955–2959 (1984).
6. Liu, F.-T., Albrandt, K. A., Bry, C. G. & Ishizaka, T. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5369–5373 (1984).
7. Coleman, J. W., Helm, B. A., Stanworth, D. R. & Gould, H. J. *Eur. J. Immun.* **15**, 966–969 (1985).
8. Geha, R. S., Helm, B. & Gould, H. *Nature* **315**, 577–578 (1985).
9. Ishizaka, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 8323–8327 (1986).
10. Padlan, E. A. & Davies, D. R. *Molec. Immun.* **23**, 1063–1075 (1986).
11. Ferguson, K. A. *Metabolism* **13**, 982–1002 (1964).
12. Prausnitz, C. & Kustner, H. *Zbl. Bkt. Orig.* **86**, 166–169 (1921).
13. Stanworth, D. R., Kings, M., Roy, P. D., Moran, J. M. & Moran, D. M. *Biochem. J.* **180**, 665–668 (1979).
14. Burt, D. S. & Stanworth, D. R. *Eur. J. Immun.* **17**, 437–440 (1987).
15. Prenner, B. M. *Annals Allergy* **55**, 332–335 (1987).
16. Ashihara, Y., Arata, Y. & Hamaguchi, K. *J. Biochem. Tokyo* **97**, 517–528 (1985).
17. Pumphrey, R. *Immun. Today* **7**, 174–178 (1986).
18. Deisenhofer, J. *Biochemistry* **20**, 2361–2370 (1981).
19. Holowka, D., Conrad, D. H. & Baird, B. *Biochemistry* **24**, 6260–6267 (1985).
20. Baenziger, J. V. in *Immediate Hypersensitivity, Modern Concepts and Developments* (ed. Bach, M. K.) 37–46 (Dekker, N.Y., 1978).
21. Kronvall, G. & Frommel, D. *Immunochimistry* **7**, 124–127 (1970).
22. Metzger, H. *et al. A. Rev. Immun.* **4**, 419–470 (1986).
23. Peterson, G. L. & Dorrington, K. J. *J. biol. Chem.* **249**, 5633–5641 (1974).
24. Laemmli, U. K. *Nature* **227**, 680–685 (1971).

Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences

James A. Lake

Molecular Biology Institute and Department of Biology,
University of California, Los Angeles, California 90024, USA

The origin of the eukaryotic nucleus is difficult to reconstruct. Eukaryotic organelles (chloroplast, mitochondrion) are eubacterial^{1,2} endosymbionts³, but the source of nuclear genes has been obscured by multiple nucleotide substitutions. Using evolutionary parsimony⁴, a newly developed rate-invariant treeing algorithm, the eukaryotic ribosomal rRNA genes are shown to have evolved from the eocytes⁵, a group of extremely thermophilic, sulphur-metabolizing, anaerobic cells. The deepest bifurcation yet found separates the reconstructed tree into two taxonomic divisions. These are a proto-eukaryotic group (karyotes) and an essentially bacterial one (parkaryotes). Within the precision of the rooting procedure, the tree is not consistent with either the prokaryotic-eukaryotic or the archaeobacterial-eubacterial-eukaryotic groupings. It implies that the last common ancestor of extant life, and the early ancestors of eukaryotes, probably lacked nuclei, metabolized sulphur and lived at near-boiling temperatures.

Reconstructing evolutionary trees that relate distant organisms presents special difficulties. The most extensively used algorithms are the parsimony and distance matrix procedures. Both types can fail when rates of nucleotide substitution are greatly unequal in juxtaposed branches, and when substitutions are frequent^{6,7}. An even more serious disadvantage is that both methods are statistically inconsistent^{6,7} in that they select the wrong tree topology more strongly when longer sequences are analysed. Unequal rate effects refer to conditions under which both methods fail (although their mathematical causes may differ). These conditions are likely to prevail when distant organisms are compared.

In a recent simulation, evolutionary parsimony was shown to be the algorithm least biased by unequal rate effects⁷. This algorithm, used to derive the tree in this paper, works by subtracting background counts (that arise from peripheral tree branches and cause unequal rate effects) from parsimony counts⁴. In contrast, the archaeobacterial tree, constructed using an iterative distance matrix algorithm⁸, is likely to be sensitive

to unequal rates. Furthermore, that tree was derived from an estimate of the number of substitutions (rather than nucleotide differences), a procedure eschewed for very distantly related sequences⁷. Recent cladistic analyses of 5S and 16S rRNAs (ref. 9) conflict with the archaeobacterial tree (and support this tree), further suggesting that the archaeobacterial tree could be an artefact of unequal rates¹⁰.

The tree reconstructed from rRNA sequences of the known types of organisms is shown in Fig. 1. It separates the eubacteria, halobacteria and methanogens from the eocytes and eukaryotes by a deep central division. The eubacteria and eukaryotes ('fast clock' organisms) exhibit substitution rates nearly ten times greater than those of their 'slow clock' sister groups. The halobacteria have an intermediate substitution rate. The juxtaposition of fast and slow clock organisms is such that it can create a large unequal rate bias.

The tree relating eubacteria, halobacteria, eocytes and eukaryotes can be joined in one of three possible topologies, corresponding to three competing theories. These are the eocyte tree (eocytes-eukaryotes, eubacteria-halobacteria), the archaeobacterial tree (eubacteria-eukaryotes) and the halobacterial tree (halobacteria-eukaryotes). As conflicting results had been reported for this part of the tree, it was analysed in detail.

Tree topologies can be affected by sequence alignments. Hence, two different alignments were tested. In the star alignment (described in the legend to Fig. 2), rRNA sequences were aligned with respect to a standard (*Thermoproteus tenax*) sequence, using a global computer algorithm. The literature alignment (see the legend to Fig. 2), used a published eubacterial-eukaryotic alignment. The significance plots in Fig. 2 indicate strong support for the eocyte tree for both alignments (between 0.002% and 0.02%), and do not support either the archaeobacterial or halobacterial trees.

As a further check against misalignment or other artefacts, the sequence was subdivided¹¹ at the midpoint into 5' and 3'-fragments. The two were analysed separately. The results, shown in Fig. 2b, indicate significant support for the eocyte tree and provide none for the two alternative subtrees. The 3' half of the RNA supports the eocyte tree more strongly, in accord with the finding that substitutions occur more frequently in the 5' half of 16S rRNA (ref. 12).

Because tree reconstruction algorithms vary in their sensitivity to the unequal rate effects, results from various methods were compared. The results are shown in Fig. 2c. Neighbour joining^{13,7} and maximum parsimony¹⁴, most sensitive to unequal rate artefacts⁷, support the archaeobacterial tree, but distribute their counts nearly equally among all three trees, suggesting the

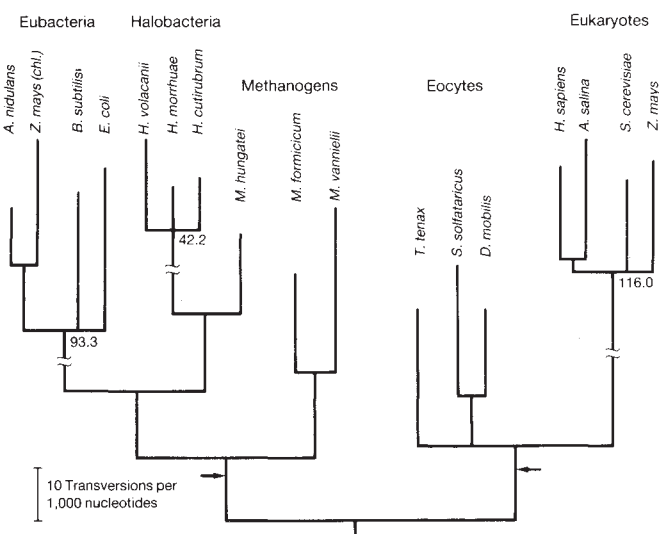
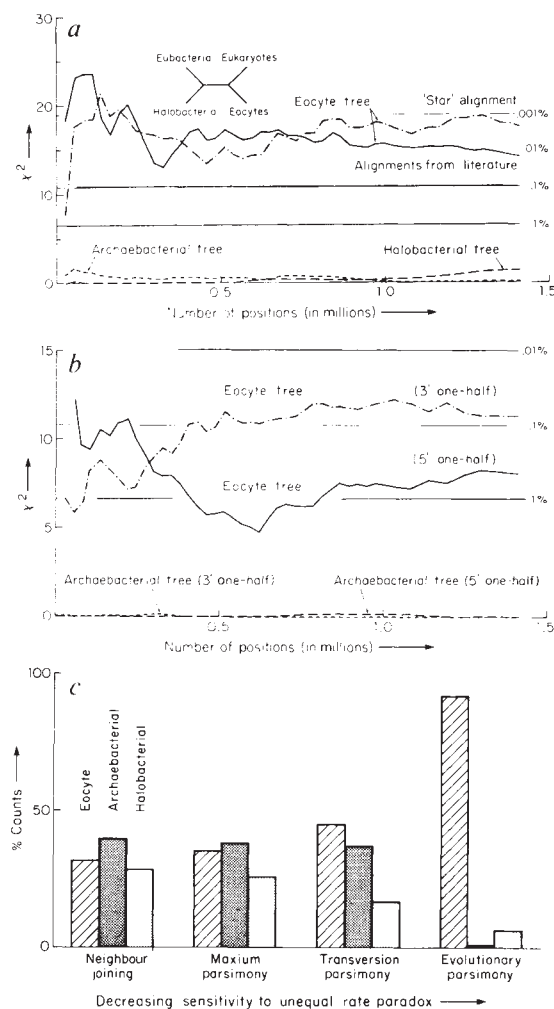


Fig. 1 The rooted evolutionary tree that relates all known groups of extant organisms. The five groups are (from left to right) eubacteria, halobacteria, methanogens, eocytes, and eukaryotes. Eubacteria and eukaryotes (and halobacteria to a lesser extent) are fast clock organisms, hence the lengths of the branches are listed. The rate-independent technique of evolutionary parsimony⁴, which estimates the length and statistical significance of the central branch of four taxa trees, was used in conjunction with the neighbourliness procedure¹³ to reconstruct the unrooted tree. Branch lengths were determined by operator metrics²⁶ and represent numbers of transversion differences rather than substitutions. Once the tree topology was determined the root was selected by parsimony rooting. This procedure is based on the simultaneous application of two criteria, namely parsimony¹⁴ and the relative rate test²⁷. The 'best' tree is the one that uses the minimum number of rate changes to satisfy the relative rate test. The most parsimonious solution corresponded to three rate changes (in the branches leading to the eubacteria, halobacteria and eukaryotes). The 5% significance level for the most parsimonious rooting is shown by the two horizontal arrows on either side of the root.

Fig. 2 Analysis of the central branch relating eubacteria, halobacteria, eocytes and eukaryotes. A total of 1,401,504 nucleotide quartets were analysed. *a*, Significance plot illustrating the effect of alternative alignments on tree selection. Alignments were constructed for 32 representative 16S rRNA sequences, from the eukaryotic (10 taxa), eocytic (3 taxa), methanogenic (3 taxa), halobacterial (4 taxa), and eubacterial (12 taxa) lineages. Sequences are listed in the compilation by Huysmans and De Wachter²⁸. Two alignment procedures were followed. In the star alignment, which is unbiased and attempts to favour no tree-branching pattern, representative members of the five groups were aligned pairwise with the eocyte *Thermoproteus tenax*, using the Align program of the Dayhoff package. Eubacterial, halobacterial, methanobacterial, eocytic and eukaryotic sequences were subsequently aligned pairwise with the representative sequence for each group. Identities, transitions, transversions and gaps were scored as 3, 1, 0 and -6, respectively. Two regions (48-367 and 1,431-1,469 in the *E. coli* numbering system) could not be aligned consistently among all groups. Hence for interkingdom comparisons, these were not included in either alignment. The literature alignment is a reconstruction from published papers using the eukaryotic-eubacterial alignment of McCarroll *et al.*²⁹ and the prokaryotic alignment of Bock and co-workers³⁰. These alignments might have been expected to bias results to favour the archaeobacterial tree. This did not occur, indicating the alternative alignments are not significantly biasing tree selection. The significance plot shows that both alignments support the eocyte tree strongly at a level between 0.01% and 0.001%. The star alignments of data, exactly as used, are available on request from the author. *b*, Significance plots for the 5' and 3' ends (subdivided between *E. coli* nucleotide positions 771-772) of small subunit rRNAs. Both parts of the sequence support the eocyte tree significantly. As observed for the entire sequence, neither one-half sequence supports either the archaeobacterial tree or the halobacterial tree (not shown). *c*, A comparison of four different algorithms and their support for the three alternative trees. In general the algorithms most sensitive to the unequal rate effects are on the left, and those on the right are least sensitive. Evolutionary parsimony at the far right is theoretically insensitive to the unequal rate effects. Total counts were as follows: neighbour joining (482,044), maximum parsimony (116,110), transversion parsimony (52,915) and evolutionary parsimony (13,411). Because the rRNAs of eocytes are GC-rich, evolutionary parsimony, like most methods of phylogenetic inference, can be biased by GC effects. It does not appear that the analysis of this branch has been extensive, however, as the two 'zero' invariants (corresponding to the archaeobacterial and halobacterial topologies) which are expected to be most sensitive to this effect, were not significantly non-zero. The largest 'zero' invariants, possibly corresponding to GC effects, were encountered during the placement of the *M. hungatei* branch. Hence its position relative to the halobacterial and eubacterial branches should be regarded as tentative.



unequal rates are biasing them. Transversion parsimony^{14,15}, which is less sensitive to the unequal rate effects, favours the eocyte tree. Evolutionary parsimony is least sensitive and exclusively supports the eocyte tree. Taken together, this indicates that the two methods most sensitive to the unequal rate effects are being biased towards an incorrect tree, and that, when these biases are properly accounted for, the eocyte tree is strongly favoured.

As biological thought can be profoundly affected by classifications, it is useful to examine past proposals. The traditional eukaryotic-prokaryotic superkingdoms group organisms according to the presence or absence of a nucleus. Prokaryotes are thus negatively defined by lack of a nucleus¹⁶. As shown in Fig. 3, the prokaryotic classification does not fit because its two groups are derived from separate ancestors (shown as filled circles). (The universal ancestor at the bottom of the tree cannot belong to any superkingdom as it is the ancestor of the higher group-life.) Thus, prokaryotes are polyphyletic with eubacteria, halobacteria, and methanogens on one side, and eocytes on the other.

Dividing prokaryotes into eubacteria and archaeobacteria¹⁷ does not eliminate polyphyly because archaeobacteria were based on an incorrect tree topology, as the data of Fig. 2 show, and on a negative definition (they are non-eubacterial prokaryotes). Like prokaryotes, archaeobacteria are polyphyletic.

The tree suggests that two superkingdoms can be defined by the deep central bifurcation, shown in Fig. 4. The taxon on the

right consists of eocytes and eukaryotes. By naming this group karyotes (Karyotae) one removes the redundancy that was previously implied in the name eukaryotes (true + karyotes) and communicates that their common ancestor contributed genes to both eocytes and the nucleus. Karyotes share features of ribosomal operon organization, rRNA introns¹⁸, ribosome structure⁵ and polymerase structure¹⁹ that are eukaryotic. The taxon on the left (eubacteria, halobacteria and methanogenic bacteria) is fundamentally a bacterial group, having such properties as eubacterial-like ribosomal RNA operons. Hence, the name parkaryotes (Parkaryotae) reflects an independent phylogenetic status on a par with that of the karyotes, and corrects the antecedent relationship implied by the term prokaryotes. This proposal generates monophyletic²⁰ and balanced²¹ superkingdoms. Both taxa are rich in biochemical diversity and both contain large numbers of species, such as eubacteria and eukaryotes.

Parsimony analysis¹³ predicts the last common ancestor of all organisms was probably sulphur-metabolizing, anaerobic and thermophilic. Parsimony infers the character state at nodes on a tree from their distribution in extant organisms. For this analysis, I identified as sulphur metabolizers the three eocytes, *Thermoproteus tenax*, *Desulfurococcus mobilis*, and *Sulfolobus solfataricus*, and four parkaryotes, *Thermoplasma acidophilus* (not shown on tree), *Methanospirillum hungatei*, *Methanobacterium formicicum* and *Methanococcus vannielli* (methanogens can reduce sulphur in the presence of hydrogen²²). From this,

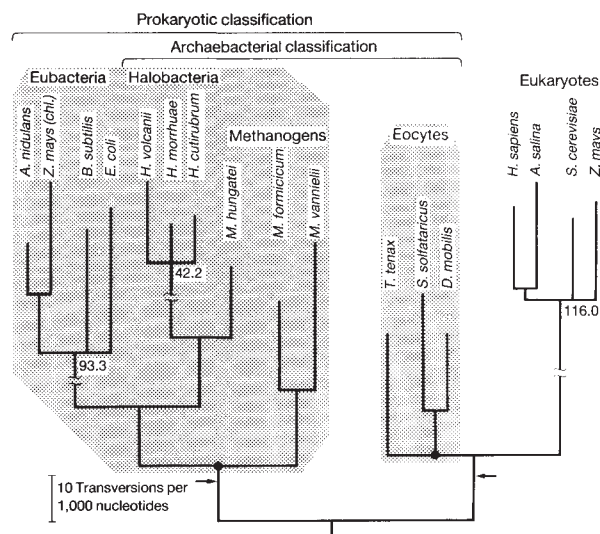


Fig. 3 Two commonly used classification schemes produce polyphyletic subgroups. In the eukaryotic-prokaryotic scheme³¹ the prokaryotes (the non-eukaryotic organisms) are polyphyletic. In the archaeobacterial-eubacterial-eukaryotic scheme¹⁷, the archaeobacteria (the non-eubacterial prokaryotes) are polyphyletic. Common ancestors of both parts of the prokaryotes are indicated by the solid circles.

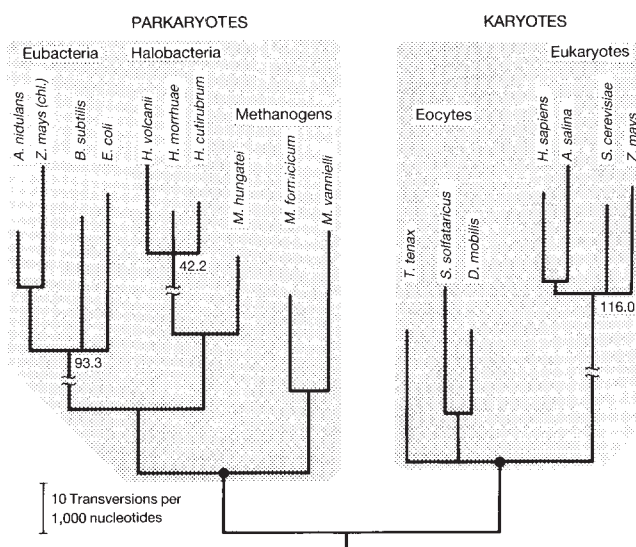


Fig. 4 The parkaryotic-karyotic classification proposed in this paper defines two balanced monophyletic groups. It organizes all extant organisms into two monophyletic superkingdoms; the parkaryotes and the karyotes. Each taxon contains a subgroup with a large number of species (eubacteria and eukaryotes) and each contains groups representing great biochemical and molecular diversity. The parkaryotes contain the eubacteria, halobacteria and methanogens (including *Thermoplasma* and *Archaeoglobus*). The karyotes consist of the eocytes and the eukaryotes. Group membership is, of course, defined phylogenetically and not metabolically; for example, the sulphur-metabolizing *Thermoplasma* is a parkaryote.

I calculated that sulphur metabolism is primitive for 98.3% of the possible trees. These preliminary analyses and related ones of thermophily suggest that the last common ancestor could have inhabited geothermal sulphur springs, either benthic or shallow, environments appropriate for the Archean²³.

At this point it is useful to discuss some details and exceptions. First, no tree can accurately represent an entire analysis. The eukaryotic divergence, for example, is presented as an unspecified trifurcation with the eocytes. In fact, in the analysis *T. tenax* is the deepest branch and the eukaryotes arose from the *D. mobilis* lineage (weakly supported, 4% significance). Because this detail is critical for understanding the origin of eukaryotes, the branching order was not specified. If eukaryotes are closest to *D. mobilis* then it would strengthen the arguments for a sulphur-metabolizing, thermophilic universal ancestor, and if not, then it could weaken them. Secondly, the selection of the root can affect the classification. For example, if the root were placed in the eukaryotic branch, this would support the prokaryotic classification. The precision of the rooting process argues against the placement (0.6% significance) and others that could alter the classification, but misrooting cannot be entirely ruled out. Finally, character distributions provide important independent data for testing the tree. Although most characters support the karyotic tree in this paper, I note two exceptions. First, the

eukaryotic rhodopsins have structural similarities with the bacteriorhodopsin of halobacteria²⁴. This argues for the halobacterial tree and against both the karyotic and archaeobacterial trees. Second, ether lipids found in halobacteria, methanogens and eocytes argue for the archaeobacterial tree and against the halobacterial and karyotic trees. If both these exceptions were valid, no tree could be correct. This suggests a need for caution in dealing with eukaryotic properties and a consideration of the chimeric, organellar origins of some nuclear DNAs. Except for these two counter examples, numerous synapomorphies representing fundamental molecular properties²⁵ support this tree.

Determining the evolutionary tree that relates all organisms has been hindered by a lack of sequence analysis techniques able to discriminate among deep branches. This is now possible, and this knowledge should bring a new and deeper comprehension of the origins of life.

This work was supported by research grants from the NSF and from the NIH, Institute of General Medical Science. I thank R. Zimmerman and R. Garrett for providing 16S rRNA sequences in advance of publication for *T. acidophilum* and *D. mobilis* respectively.

Received 28 September; accepted 26 October 1987.

- Schwartz, Z. & Kossel, H. *Nature* **238**, 739-772 (1980).
- Spencer, D. F., Schnare, M. N. & Gray, M. W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 493-497 (1984).
- Margulis, L. *Symbiosis in Cell Evolution* (Freeman, San Francisco, 1981).
- Lake, J. A. *Molec. biol. Evol.* **4**, 167 (1987).
- Lake, J. A., Hendersen, E., Oakes, M. & Clark, M. W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3786-3790 (1984).
- Felsenstein, J. *Syst. Zool.* **27**, 401-410 (1978).
- Li, W. H., Wolfe, K. H., Soudis, J. & Sharp, P. M. *Symp. Quant. Biol.* **52** (in the press).
- Pace, N. R., Stahl, D. A., Lane, D. J. & Olsen, G. J. *Adv. microbiol. Ecol.* **1**-55 (1986).
- Erdmann, V. A. et al. in *Endocytobiology III* (eds J. J. Lee & J. F. Frederick) (New York Academy of Sciences, 1987).
- Lake, J. A. *Nature* **319**, 626 (1986).
- Penny, D. & Hendy, M. D. *Cladistics* **1**(3), 266-278 (1981).
- Hickson, J. E. & Brown, W. M. *Molec. biol. Evol.* **3**, 1-18 (1986).
- Fitch, W. M. *J. molec. Evol.* **18**, 30-37 (1981).
- Fitch, W. *Am. Nat.* **111**, 223-257 (1977).

- Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. *J. molec. Evol.* **18**, 225-239 (1982).
- Eldridge, N. & Cracraft, J. *Phylogenetic Patterns and the Evolutionary Process* (Columbia University Press, New York, 1980).
- Fox, G. E. et al. *Science* **209**, 457-463 (1980).
- Larsen, N., Leffers, H., Kjems, J. & Garrett, R. A. *System. Appl. Microbiol.* **7**, 49-57 (1986).
- Zillig, W., Schnabel, R. & Stetter, K. O. *Curr. Top. Microbiol. Immun.* **33**, 1-18 (1985).
- Wiley, E. O. *Phylogenetics* (Wiley, New York, 1981).
- Mayr, E. *Principles of Systematic Zoology* (McGraw-Hill, New York, 1969).
- Stetter, K. O. & Gaag, G. *Nature* **305**, 309-311 (1983).
- Schopf, J. W. *Earth's Earliest Biosphere* (Princeton University Press, New Jersey, 1983).
- Findlay, J. B. C. & Pappin, D. J. C. *Biochem. J.* **238**, 625-642 (1986).
- Lake, J. A. *Nature* **321**, 657-658 (1986).
- Lake, J. A. *J. molec. Evol.* **26**, 59-73 (1987).
- Wilson, A., Carlson, S. & White, T. A. *Rev. Biochem.* **46**, 573-639 (1977).
- Huysmans, E. & DeWachter, R. *Nucleic Acids Res.* **14**, 73-118 (1986).
- McCarroll, R., Olsen, G. J., Stahl, Y. D., Woese, C. R. & Soglin, M. S. *Biochemistry* **22**, 5858-5868 (1983).
- Leinfelder, W., Jarsch, M. & Bock, A. *System. Appl. Microbiol.* **6**, 164-170 (1985).
- Allsopp, A. *New Phytol.* **68**, 591-612 (1969).

† To whom correspondence should be addressed.

Table 1 Aminoacylation with glutamate and glutamine of different tRNA samples with various preparations of aminoacyl-tRNA synthetase.

tRNA S100	Barley (chloro)		Spinach (chloro) total	Mung bean (total)	Mouse (mito)	Synecho- cystis	<i>B. subtilis</i> *		<i>E. coli</i> †		Barley (Cyto) Glu	
	DALA	Gln					Glu	Gln	Glu	Gln		
Barley (chloro)	+	+					+	—	+	—	—	Glu
	—	—					—	—	—	—		Gln
Spinach (chloro)			+						+			Glu
			—						—			Gln
Mung bean (mito)	+	+		+					+			Glu
	—	—		—					—			Gln
Mouse (mito)	+	+			+				+			Glu
	—	—			—				—			Gln
<i>Synechocystis</i>	+	+				+			+	—		Glu
	—	—				—			—	—		Gln
<i>B. subtilis</i>	—	+					+	+	—	(+) [†]		Glu
	—	—					—	—	—	—		Gln
<i>E. coli</i>	+	—					—	—	+	—		Glu
	—	—					—	+	—	+		Gln

Test for the presence of direct Gln-tRNA^{Gln} formation in organelles and some prokaryotic organisms. Aminoacyl-tRNA synthetases and tRNAs from the sources described were used to charge tRNA with glutamate or glutamine. To exclude the possibility of inhibition of GlnRS by unknown compounds in the extract, we used a partially purified synthetase preparation (Blue-Sepharose fraction, ref. 18) in addition to a crude barley chloroplast S-100 preparation. Barley plastids (from plants grown in the dark for 6 days and greened under white fluorescent light for 20 h) and spinach chloroplasts (from commercially grown material) were prepared according to ref. 24. Mung-bean mitochondria were prepared from commercially grown material according to ref. 25. All plant organelles were subjected to final purification on a Percoll-step gradient. Mouse liver mitochondria were isolated according to ref. 26. Bacteria were harvested in mid-log phase. Crude synthetases were prepared by lysing the organelles in five volumes of synthetase buffer (50 mM Tris-HCl, pH 7.5; 1 mM Na₂EDTA; 10 mM MgCl₂; 20 mM KCl; 10 mM β-mercaptoethanol) containing 0.1% Triton X-100 or by grinding the bacteria with sea sand in a small volume of synthetase buffer and separating the lysate from the sand by low-speed centrifugation. Ribosomes and membrane fragments were removed by centrifugation at 100,000g for 1 h, the supernatant was dialysed extensively against synthetase buffer containing 50% glycerol and stored at -20 °C. tRNAs from organelles or bacteria were prepared according to ref. 27. Charging assays were performed in a total volume of 15 μl as described in the legend to Fig. 3, using either 13.3 A₂₆₀ per ml of HPLC-enriched barley chloroplast tRNAs or 33 A₂₆₀ per ml of the respective total tRNA. Incubations were for 20 min at 28 °C for extracts from plant organelles and *Synechocystis*, 5 min at 37 °C for mouse mitochondria and 10 min at 37 °C for *B. subtilis* and *E. coli*.

* The discrimination between tRNA^{Gln} and tRNA^{Glu} from *B. subtilis* was made by precharging a total tRNA preparation with [¹²C]glutamine by *E. coli* synthetases, which charge only tRNA^{Gln}, but not tRNA^{Glu} from *B. subtilis*. This precharged tRNA was used in a charging assay with the barley Blue-Sepharose fraction and [¹⁴C]glutamate.

† It has been reported⁹ that *E. coli* tRNA^{Gln} can be glutamylated by *B. subtilis* GluRS; but we were not able to reproduce this result under our standard charging conditions.

be glutamylated by a crude *Escherichia coli* synthetase preparation which glutamylates tRNA^{Glu} albeit at a lower rate (Fig. 3). On the other hand, a crude *Bacillus subtilis* cell extract charges the tRNA^{Gln} species, but not tRNA^{Glu}, with glutamate. Thus, all three barley chloroplast glutamate-accepting tRNA species are recognized efficiently by homologous or heterologous glutamyl-tRNA synthetases. But they cannot be glutaminylated by *E. coli* aminoacyl-tRNA synthetases or by homologous extracts (Table 1). Even when total chloroplast tRNA is used as a substrate, glutaminyl-tRNA synthetase (GlnRS) activity cannot be detected in barley chloroplast extracts. Similarly, these preparations are unable to charge unfractionated *B. subtilis* and *E. coli* tRNA or pure *E. coli* tRNA^{Glu} and tRNA^{Gln} species with glutamine (Table 1). The possibility that these observations arise from the presence of an inhibitor of GlnRS in the barley chloroplast extract was eliminated by performing a mixing experiment: partially purified *E. coli* tRNA^{Gln} can be charged by *E. coli* GlnRS to the same extent (~900 pmoles per A₂₆₀ unit) in the presence or absence of barley extract. We thus conclude that there is no detectable GlnRS activity present in barley chloroplasts.

Is the lack of GlnRS a more general phenomenon? Knowing that some Gram-positive organisms also lack GlnRS⁵, we decided to check other prokaryotic organisms and organelles from other species for the presence or absence of this enzyme. These charging experiments were performed with unfractionated aminoacyl-tRNA synthetase preparations from spinach chloroplasts, mung-bean mitochondria, mouse mitochondria and the cyanobacterium *Synechocystis* 6803, using homologous unfractionated tRNAs and also unfractionated tRNA or pure tRNA^{Glu}

and tRNA^{Gln} species from *E. coli*. The experimental values were compared to the values obtained with a *B. subtilis* aminoacyl-tRNA synthetase preparation which lacks GlnRS^{5,6}. As Table 1 shows, there is no GlnRS activity present in any of the different aminoacyl-tRNA synthetase preparations except in the one from *E. coli*. It has been reported that bean chloroplast tRNA can be glutaminylated by a homologous enzyme fraction⁷. Although these data suggest the presence of GlnRS in the organelle, the nature of the radioactive amino acid actually attached to tRNA was not determined, even though a moderate excess of glutamate was used in the experiment. Furthermore, a glutamine-accepting tRNA has been identified in *Euglena* chloroplasts⁸; however the source of synthetase and the reaction conditions have not been indicated. Thus, further experiments are needed to determine whether GlnRS is really present in these cases. Considering all these observations we suggest that there is no detectable GlnRS in organelles of plant and animal origin and in cyanobacteria.

The unexpected finding that the two barley chloroplast glutamate isoacceptors, incapable of forming δ-aminolevulinatate, are in fact tRNA^{Gln} species established that the tRNA^{Glu} species in this organelle plays a dual role: its glutamate is used in protein biosynthesis and it also provides the activated substrate for the dehydrogenase involved in δ-aminolevulinatate biosynthesis¹.

The pathway of Gln-tRNA^{Gln} formation in chloroplasts is very similar to the one described for bacilli²: tRNA^{Gln} is misacylated with glutamate by a glutamyl-tRNA synthetase (GluRS)⁶, presumably the same enzyme that charges tRNA^{Gln}. In the presence of a suitable amide donor, Glu-tRNA^{Gln} is then converted to Gln-tRNA^{Gln} by an amidotransferase. Thus, two enzymes

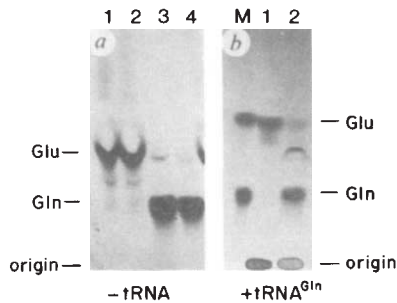


Fig. 2 Conversion of Glu-tRNA^{Gln} to Gln-tRNA^{Gln} by a tRNA-dependent amidotransferase. *a*, Incubation of glutamate and glutamine in barley chloroplast extract (Blue-Sepharose fraction, ref. 18) under aminoacylation conditions (see legend to Fig. 3) but without added tRNA. Amino acids were added as follows: Lane 1, [¹⁴C]Glu: 25 μ M [¹⁴C]glutamate only; lane 2, [¹⁴C]Glu + [¹²C]Gln: same but with addition of 1.25 mM [¹²C]glutamine; lane 3, [¹⁴C]Gln: 25 μ M [¹⁴C]glutamine; lane 4, [¹⁴C]Gln + [¹²C]Glu, same but with addition of 1.25 mM [¹²C]glutamate. *b*, Conversion of Glu-tRNA^{Gln} to Gln-tRNA^{Gln} in the presence of an amide donor. Reaction conditions were as described in panel *a*, except that partially purified tRNA^{Gln} (13.3 A₂₆₀ per ml of a tRNA^{Gln}-containing HPLC-fraction) was added. Lane M, a mixture of [¹⁴C]Gln and [¹⁴C]Glu used as reference markers. Lane 1, [¹⁴C]Glu, aminoacylation of tRNA^{Gln} with [¹⁴C]glutamate. Lane 2, [¹⁴C]Glu + [¹²C]Gln, aminoacylation of tRNA^{Gln} with [¹⁴C]glutamate in presence of 1.25 mM [¹²C]glutamine.

Methods. The reactions were carried out for 20 min at 28 °C in a total volume of 15 μ l and stopped by phenol extraction. *a*, The aqueous phases were dried down and analysed by cellulose TLC in isopropanol:formic acid:water = 80:20:4 (ref. 23). *b*, The extracted aminoacyl-tRNAs, ethanol precipitated, and the amino acids released by incubation with 10 mM Tris-HCl (pH 9) for 30 min at 37 °C. The dried samples were analysed as described for panel *a*.

are required for the synthesis of Gln-tRNA^{Gln}: a mischarging GluRS and a tRNA-dependent amidotransferase. The amidotransferase is formally similar to GlnRS; both enzymes possess specific binding sites for tRNA^{Gln} and for glutamine. The mechanism by which the cell copes with the mischarging phenomenon while it maintains the fidelity in protein synthesis, serving the glutamine codons found in the chloroplast genome is intriguing. Clearly, the mischarged tRNA must not be present in the free pool of aminoacyl-tRNAs in the cell. Possibly the ribosome has the ability to discriminate against this tRNA, or the affinity of the translational elongation factor Tu (EF-Tu) is significantly lower for Glu-tRNA^{Gln} and higher for Gln-tRNA^{Gln}. This would require a subtle change in tRNA conformation upon conversion of the attached amino acid. It is known that the affinity of *E. coli* EF-Tu is strongest for Gln-tRNA^{Gln} and much weaker for Glu-tRNA^{Glu} (ref. 9). But the individual contributions of the amino acid and of the tRNA to the binding strength are not known. Alternatively, there could be a number of mechanisms to ensure that this tRNA would not be available in free form. This might be achieved by channelling¹⁰ which would rely on a tight complex of GluRS and the amidotransferase; the mischarged tRNA would then be directly transferred to the amidotransferase and only released if charged with the cognate amino acid. One may also wonder whether the amidotransferase and GluRS are separate proteins, as they could possibly be encoded in the same polypeptide. There is precedence for this phenomenon: in some organisms the thymidylate synthase and tetrahydrofolate reductase, enzymes needed for sequential reactions in thymidylate formation, are encoded in separate proteins, whereas in other species both functions are encoded in the same polypeptide¹¹.

The occurrence of the pathway of Gln-tRNA^{Gln} formation by transamidation is widespread. Evidence exists for its presence in Gram-positive eubacteria^{2,5}, *Halobacterium* species^{12,13}, *Synechocystis* and plant and animal organelles (this study). Also

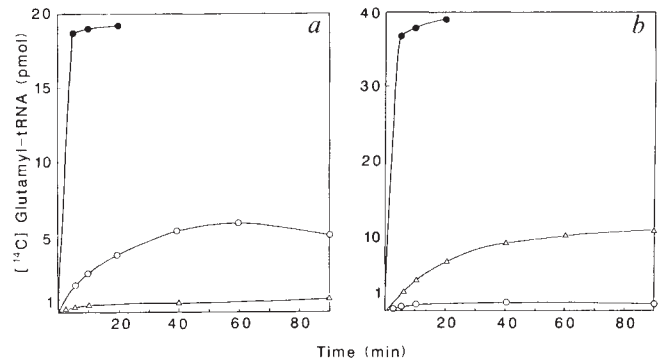


Fig. 3 Time course of glutamylation of barley chloroplast tRNA^{Glu} (panel *a*) and tRNA^{Gln} (panel *b*) with homologous and heterologous synthetases. Aminoacylation reactions contained 100 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 4 mM β -mercaptoethanol, 5 mM ATP, 13.3 A₂₆₀ per ml HPLC enriched barley chloroplast tRNA, 25 μ M [¹⁴C]glutamate (280 mCi per mmol), and 0.2–0.6 mg per ml protein. Incubation was at 28 °C for the chloroplast enzymes and at 37 °C for the bacterial enzymes. Aliquots of 15 μ l were spotted onto paper filter disks, washed in ice-cold 10% and 5% trichloroacetic (TCA), rinsed in ethanol, dried and counted for radioactivity after the addition of scintillation fluid. The tRNA^{Glu} and tRNA^{Gln} fractions were 12 and 25% pure, respectively. Closed circles, barley chloroplast Blue-Sepharose fraction; open circles, *E. coli* S-100; triangles, *B. subtilis* S-100.

there is evidence to suggest that in yeast mitochondria (ref. 14; R. Martin and G. Dirheimer, unpublished data) the same situation obtains. On the other hand the direct pathway of Gln-tRNA^{Gln} formation occurs in the cytoplasm of eukaryotic cells and in Gram-negative eubacteria. It can be concluded that the mischarging mechanism of Gln-tRNA^{Gln} formation may have originated very early in evolution. Alternatively, convergent evolution at a later time may be responsible for its widespread occurrence, although it is not immediately apparent what advantage this pathway would confer. It is generally believed that eukaryotic organelles arose from endosymbiotic prokaryote-like ancestors: phylogenetic studies show a close relationship between chloroplasts and cyanobacteria or prochlorophyta, and between mitochondria and the α -subdivision of purple bacteria^{15,16}. The plant and fungal mitochondria appear to be derived from different endosymbiotic events than the animal organelle¹⁷.

It is tempting to speculate that the mischarging/transamidation pathway was already present before the presumed common ancestors of modern prokaryotes and organelles diverged, and that the GlnRS of eukaryotic cytoplasm and Gram-negative bacteria evolved independently, possibly to assure a greater degree of fidelity in protein synthesis. Our studies are in agreement with these conclusions. The conservation of the transamidation pathway of Gln-tRNA^{Gln} formation in cyanobacteria, chloroplasts and mitochondria add evidence from intermediary metabolism to the information on evolutionary relationships gained from the examination of tRNA and protein sequence analysis^{16,17}.

We thank Gary O'Neill for *Synechocystis* and *B. subtilis* extract and John Perona for *E. coli* tRNA^{Gln}. We are grateful to Peter Ray for help with preparation of mouse mitochondria. This work was supported by grants from the NIH, the NSF and the Department of Energy.

Received 29 July; accepted 4 November 1987.

- Schön, A. *et al. Nature* **322**, 281–284 (1986).
- Wilcox, M. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **61**, 229–236 (1968).
- Sprinzl, M., Hartmann, T., Meissner, F., Moll, J. & Vorderwülbecke, T. *Nucleic Acids Res.* **15**, r53–r188 (1987).
- Ohya, K. *et al. Nature* **322**, 572–574 (1986).
- Wilcox, M. *Eur. J. Biochem.* **11**, 405–412 (1969).

6. Lapointe, J., Duplain, L. & Proulx, M. *J. Bact.* **165**, 88-93 (1986).
7. Burkard, G., Guillemot, P. & Weil, J. H. *Biochim. biophys. Acta* **224**, 184-198 (1970).
8. Mubumbila, M. *et al.* *Biochim. biophys. Acta* **609**, 31-39 (1980).
9. Louie, A., Ribeiro, S., Reid, B. R. & Jurnak, F. *J. biol. Chem.* **259**, 5010-5016 (1984).
10. Srivastava, D. K. & Bernhard, S. A. *Science* **234**, 1081-1086 (1986).
11. Grumont, R., Washtien, W. L., Caput, D. & Santi, D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5387-5391 (1986).
12. White, B. N. & Bayley, S. T. *Can. J. Biochem.* **50**, 601-609 (1971).
13. Gupta, R. *J. biol. Chem.* **259**, 9461-9471 (1985).
14. Martin, N. C., Rabinowitz, M. & Fukuhara, H. *J. molec. Biol.* **101**, 285-296 (1976).
15. Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. & Woese, C. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4443-4447 (1985).
16. Woese, C. R. *Microbiol. Rev.* **61**, 221-227 (1987).
17. Gray, M. W., Sankoff, D. & Cedergren, R. *J. Nucleic Acids Res.* **12**, 5837-5852 (1984).
18. Kannangara, C. G., Gough, S. P., Oliver, R. P. & Rasmussen, S. K. *Carlsberg Res. Commun.* **49**, 417-437 (1984).
19. Krupp, G. & Gross, H. J. in *The Modified Nucleosides in Transfer RNA II: A Laboratory Manual of Genetic Analysis, Identification and Sequence Determination* (eds Agris, P. F. & Kopper, R. A.) 11-58 (Liss, New York, 1983).
20. Peattie, D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1760-1764 (1979).
21. Stanley, J. & Vassilenko, S. *Nature* **274**, 87-89 (1978).
22. Nishimura, S. in *Transfer RNA: Structure, Properties and Recognition*. (eds Schimmel, P., Söll, D. & Abelson, J. N.) 551-552 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1979).
23. von Arx, E. & Neher, R. *J. Chromatogr.* **12**, 329-341 (1963).
24. Höinghaus, R. & Feierabend, J. *Protoplasma* **118**, 114-120 (1983).
25. Jackson, D., Dench, J. E., Hall, D. O. & Moore, A. L. *Pl. Physiol.* **64**, 150-153 (1979).
26. Ledwith, B. J., Manam, S. & Van Tuyle, G. C. *J. biol. Chem.* **261**, 6571-6577 (1986).
27. Roe, B. A. *Nucleic Acids Res.* **2**, 21-42 (1975).

Drosophila small cytoplasmic 19S ribonucleoprotein is homologous to the rat multicatalytic proteinase

Patricia-E. Falkenburg, Christian Haass,
Peter-M. Kloetzel*, Barbara Niedel†,
Friedrich Kopp†, Lothar Kuehn†
& Burkhardt Dahlmann†

ZMBH/Institut für molekulare Genetik, Im Neuenheimer Feld 282,
University of Heidelberg, D-6900 Heidelberg, FRG
† Biochemische Abteilung, Diabetes Forschungsinstitut,
Auf'm Hennekamp 65, D-4000 Düsseldorf, FRG

All eukaryotic cells so far analysed contain 19S particles which share a cylinder-like shape and are composed of a set of proteins of relative molecular mass ranging typically from 19,000 to 36,000 (refs 1-10). Proposed functions have included synthetase activity¹¹, transfer RNA processing¹² or messenger RNA repression⁶, but their biological importance remains obscure. A multicatalytic proteinase (MCP) of similar size and shape has been isolated from mammalian tissues¹³⁻²⁴. The apparent similarities of these high molecular weight complexes suggest a biochemical and functional homology between the small cytoplasmic 19S particle from *Drosophila melanogaster* (19S-scRNP) (ref. 7) and rat MCP (ref. 14). By means of electron microscopy, immunological techniques, RNA identification and proteinase activity assays, we were able to show that the two structurally similar complexes are immunologically related ribonucleoproteins (RNPs) with similar proteolytic activity.

Electron micrographs of the two complexes show that the particles have the same cylinder-like shape and are composed of four rings or disks (Fig. 1*a, b* and refs 7, 25). As the polypeptide pattern of both complexes is similar, but not identical, with respect to molecular weight range and complexity (Fig. 2*a*), we analysed the immunological cross-reactivity of the proteins using antibodies directed against either the highly purified 19S-scRNP (Dm19sc, Fig. 2*b*, lane 1) or the MCP (anti-MCP, Fig. 2*b*, lane 3) (ref. 14). When antiserum Dm19sc is probed on rat liver MCP it reacts strongly with a protein of relative molecular mass (M_r) 28,500 (28.5 K) and more weakly with several smaller polypeptides of the MCP (Fig. 2*b*, lane 4). Anti-MCP antiserum identifies the 35K, 27.5K and 25K proteins of *Drosophila* 19S-scRNP

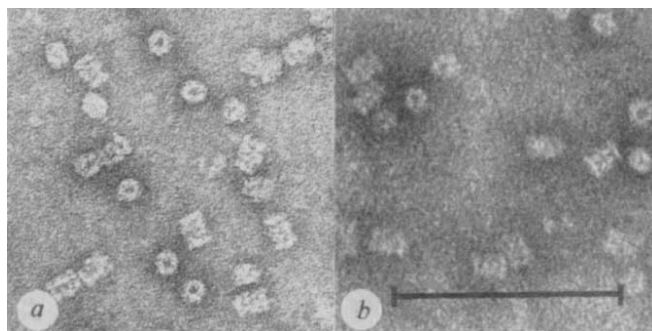


Fig. 1 Electron micrographs of *a*, 19S-scRNPs from *D. melanogaster* pupae and *b*, rat liver MCP.

Methods. Rat liver MCP was purified as previously described for skeletal muscle¹⁴. The cytoplasmic supernatant of *Drosophila* embryos was prepared and centrifuged as before⁷. 19S gradient fractions were separated on a DEAE-Sephacel column. Bound material was eluted from the column stepwise, using buffer A (200 mM potassium acetate, 10 mM HEPES pH 7.2, 5 mM magnesium acetate) for washing the column and buffer B (300 mM potassium acetate, 10 mM HEPES pH 7.2, 5 mM magnesium acetate) for eluting assayed 19S particles. Samples were negatively stained with uranyl acetate (Fig. 1*a*) or with phosphotungstate (Fig. 1*b*). Two projections of the cylinder-shaped complexes are shown as rings (~11 nm in diameter) with a dense core (~4 nm in diameter) and as a side-view (~16 nm in length). Scale bar represents 100 nm.

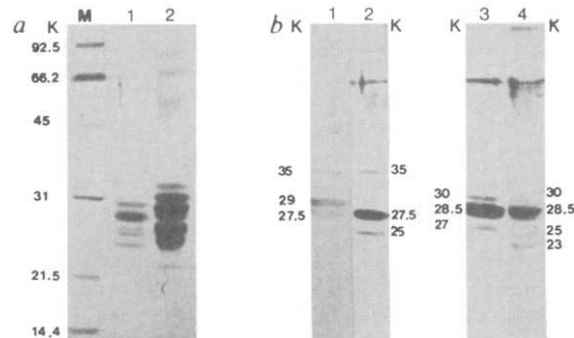


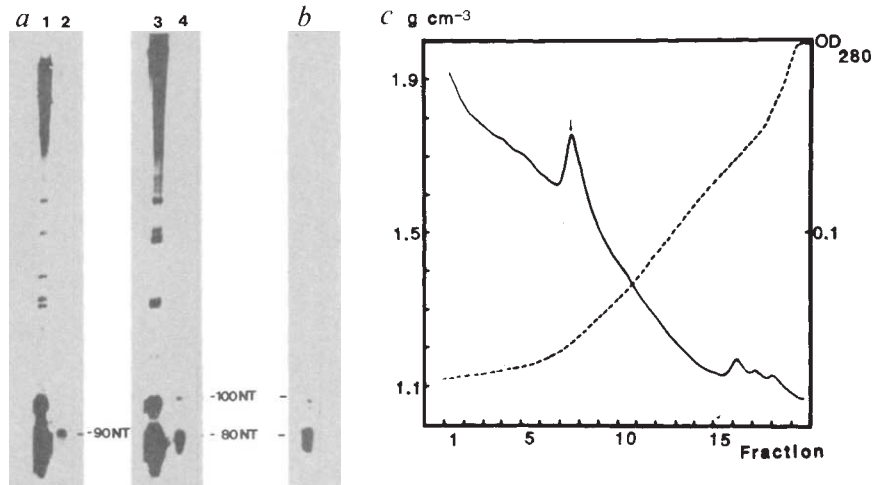
Fig. 2 Immunological cross-reactivity between *D. melanogaster* 19S-scRNP and rat liver MCP. *a*, Ponceau-staining: lane M, markers; lane 1, *Drosophila* 19S-scRNP; lane 2, rat MCP. *b*, Immunoblots: lane 1, *Drosophila* 19S-scRNP with Dm19sc; lane 2, *Drosophila* 19S-scRNP with anti-MCP; lane 3, rat MCP with anti-MCP; lane 4, rat MCP with Dm19sc. The relative molecular mass is shown (K).

Methods. Proteins were separated on SDS-polyacrylamide gels²⁸, electrophoretically transferred to nitrocellulose filters, stained with 10% Ponceau-red and probed with antibodies raised either against the *D. melanogaster* 19S-scRNP (Dm19sc) or against rat muscle MCP (anti-MCP). Immunoblotting was performed as described⁷, except that blots with heterologous antibodies were washed with 0.3 M NaCl, 10 mM Tris/HCl pH 7.2, 0.3% Triton X-100.

(Fig. 2*b*, lane 2). Previously we have shown by density gradient centrifugation and ultraviolet cross-linking experiments that the 19S-particles of *Drosophila* have properties indicative of RNP-complexes⁷. To define the RNA component, 19S particles were immunoprecipitated with Dm19sc from the soluble cytoplasmic supernatant of ³²P-labelled embryonic *Drosophila* S3 culture cells. The precipitated, *in vivo*-labelled RNA was then analysed on 7.5% polyacrylamide sequencing gels. A single RNA species ~90 nucleotides (nt) long is identified as the RNA-component of the 19S particles (Fig. 3*a*, lane 2).

* To whom correspondence should be addressed.

Fig. 3 *a*, Gel autoradiograph of RNAs. Lane 1, total RNA of the labelled fraction from *Drosophila* S3 cells; lane 2, RNA content of *Drosophila* 19S-scRNP after immunoprecipitation with Dm19sc; lane 3, total RNA of the labelled fraction from rat hepatoma cells; and lane 4, RNA content of rat MCP after immunoprecipitation. *b*, RNA-content of rat MCP purified on a Cs_2SO_4 -density gradient and *c*, buoyant density of rat liver MCP. The arrow indicates the MCP-complex which sediments at a density of $\rho = 1.23 \text{ g cm}^{-3}$. Broken line gives absorbance at 254 nm.



Methods. *D. melanogaster* S3 tissue culture cells and rat hepatoma cells (FTO2B) (ref. 26) were grown in the presence of [^{32}P]orthophosphate for 24 h. The cytoplasmic supernatant of the cells was then isolated⁷. The 19S cylinder particles were immunoprecipitated by incubating the extract in the presence of RNasin (35 U per 0.5 ml) with proteinA-Sepharose-bound antibodies Dm19sc or anti-MCP. After 1 h at 4°C the proteinA-Sepharose was washed 10 times with 0.1% Triton X-100 in 0.5 M potassium acetate, 10 mM HEPES pH 7.2, 5 mM magnesium acetate. The precipitated ^{32}P -labelled RNA was isolated by phenol extraction and analysed on 7.5% polyacrylamide sequencing gels. RNP density in Cs_2SO_4 gradients was determined as before²⁹. RNA was extracted from the density gradient fractions with phenol, 3' end-labelled with [^{32}P]pCp and T4 RNA ligase, and analysed on 7.5% polyacrylamide sequencing gels.

We investigated whether the rat liver MCP also contains RNA: a buoyant density of $\rho = 1.23 \text{ g cm}^{-3}$ is obtained by Cs_2SO_4 density gradient centrifugation, indicating a relatively high protein to RNA ratio (Fig. 3*b*). This density is slightly lower than that measured for the 19S-scRNP of *Drosophila* but still higher than that observed for proteins under the same conditions⁷. To define the RNA component of the rat MCP complex, the proteinase was immunoprecipitated from the soluble cytoplasmic supernatant of ^{32}P -labelled rat hepatoma cells (cell line FTO2B, ref. 26) and the labelled RNA analysed electrophoretically. RNA molecules in the size range of 80–100 nt are identified as the RNA component of the rat liver MCP (Fig. 3*a*, lane 4). Analysis of 3' ^{32}P -labelled RNA extracted from the MCP after density gradient centrifugation gave a similar RNA pattern (Fig. 3*b*).

Purified *Drosophila* 19S-scRNP was tested for its hydrolytic activity towards peptide substrates known to be susceptible to

MCP (ref. 14). As summarized in Table 1, the synthetic peptide substrates as well as the ^{14}C -methylcasein are split by the 19S-scRNP at specific activities similar to the MCP, the slight differences being explained by tissue or species specificities. Finally, caseinolytic activity of the 19S-scRNP is inhibited by both the antiserum raised against the complex from *Drosophila* and the antiserum to rat MCP (Table 1).

Our data demonstrate the homology between the evolutionary ubiquitous 19S particles, namely 'prosome'^{6,8–10}, 'miniparticles'^{2,3}, 'cylinder particles'^{4,5} and 'ring-type particles'⁷, and a previously described, multicatalytic proteinase^{13–24}. Comparison of the 19S particles of *Drosophila* and rat liver MCP on the basis of their morphology, molecular composition, immunological cross-reactivity, as well as common enzymatic activity indicate that these two types of particles are homologous.

The protein composition of the 19S cylinder-shaped particles of both species is not identical (Fig. 2). This could be explained by the evolutionary distance of the two species, in agreement with a recent analysis of evolutionary variation of the 19S-scRNP proteins⁹. Despite these differences, immunoblotting shows that both particles share antigenic determinants, indicating that certain epitopes have been conserved during evolution. Thus anti-MCP recognizes the 35K *Drosophila* 19S particle protein even though rat MCP contains no protein of similar size. Moreover, our experiments show that both complexes contain small RNA molecules with 80–100 nt, the difference in size and number of RNA species between rat and *Drosophila* probably being due to interspecies variation^{8–10}.

This is the first demonstration that a small cytoplasmic RNP can catalyse proteolytic reactions. It is interesting that in *Drosophila* the 19S-scRNPs are partly found in tight association with polyribosomes²⁷, and that there are development-specific variations of the complex in the two-dimensional protein pattern (C.H. and P.-M.K., in preparation). One possible function of the MCP could therefore be an involvement in development-specific processing and post-translational modification of newly synthesized proteins.

We thank Professor Dr E. K. F. Bautz for excellent facilities, and Professor Dr U. Scheer for help with electron microscope analysis of *Drosophila* 19S-scRNPs. This work was supported in part by the Deutsche Forschungsgemeinschaft and the Ministerium für Wissenschaft und Forschung des Landes Nordrhein-Westfalen, and by the Bundesministerium für Jugend, Familie, Frauen und Gesundheit.

Table 1 Multicatalytic proteinase activity of the *D. melanogaster* 19S-scRNP and rat liver MCP

Substrate	Specific activity (pmol min^{-1} per mg protein)	
	MCP	19S-scRNP
Bz-Val-Gly-Arg-NMec	3,570	2,022
Suc-Ala-Ala-Phe-NMec	496	282
Suc-Leu-Leu-Val-Tyr-NMec	612	1,210
	(c.p.m. per μg protein)	
^{14}C -methylcasein	713	1,385 (100%)
After preincubation with antibodies:		
Dm19sc	n.d.	398 (29%)
anti-MCP	n.d.	348 (25%)

Determination of proteolytic activities of the purified complexes towards peptide-NMec (4-methyl-7-coumarylamide) substrates as well as ^{14}C -methylcasein were performed as described¹⁴. Time of incubation with ^{14}C -methylcasein was 60 min. Immuno-inhibition assays of 19S-scRNP caseinolytic activity were done with 20-fold (Dm19sc) and 40-fold (anti-MCP) excess of IgG protein. After preincubation for 60 min at 21°C, samples were centrifuged for 10 min (45,000g) and caseinolytic activity in the supernatant was measured. Values in brackets correspond to percent activity as compared to the value without added antibodies (n.d., not determined).

Received 31 July; accepted 2 November 1987.

- Kloetzel, P.-M. *Molec. Biol. Rep.* **12**, 223-227 (1987).
- Domae, N. *et al. Life Sci.* **30**, 469-477 (1981).
- Harmon, F. R., Spohn, W. H., Domae, N., Soo Ha, C. & Busch, H. *Cell Biol. int. Rep.* **7**, 333-343 (1983).
- Kleinschmidt, J. A., Hügler, B., Grund, C. & Franke, W. W. *Eur. J. cell Biol.* **32**, 143-156 (1983).
- Hügler, B., Kleinschmidt, J. A. & Franke, W. W. *Eur. J. cell Biol.* **32**, 157-163 (1983).
- Schmid, H.-P. *et al. EMBO J.* **3**, 29-34 (1984).
- Schuldt, C. & Kloetzel, P.-M. *Dev. Biol.* **110**, 65-74 (1985).
- Arrigo, A.-P., Darlix, J.-L., Khandjian, E. W., Simon, M. & Spahr, P.-F. *EMBO J.* **4**, 399-406 (1985).
- Arrigo, A.-P. *J. molec. Evol.* **25**, 141-150 (1987).
- de Sa, C. M. *J. molec. Biol.* **187**, 479-493 (1986).
- Shelton, E., Kuff, E. L., Maxwell, E. S. & Harrington, J. T. *J. cell Biol.* **45**, 1-8 (1970).
- Castaño, J., Ornberg, R., Koster, J. G., Tobian, J. A. & Zasloff, M. *Cell* **46**, 377-387 (1986).
- Wilk, S. & Orlowski, M. *J. Neurochem.* **40**, 842-849 (1983).
- Dahlmann, B., Kuehn, L., Rutschmann, M. & Reinauer, H. *Biochem. J.* **228**, 161-170 (1985).
- Ray, K. & Harris, H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7545-7549 (1985).
- Rivett, A. J. *J. biol. Chem.* **260**, 12600-12606 (1985).
- Nojima, M. *et al. J. Biochem., Tokyo* **99**, 1605-1611 (1986).
- Ishiura, S. & Sugita, H. *J. Biochem., Tokyo* **100**, 753-763 (1986).
- McGuire, M. J. & De Martino, G. N. *Biochem. biophys. Acta* **873**, 279-289 (1986).
- Yamamoto, T., Nojima, M., Ishiura, S. & Sugita, H. *Biochem. biophys. Acta* **882**, 297-304 (1986).
- Tanaka, K., Ii, K., Ichihara, A., Waxman, L. & Goldberg, A. L. *J. biol. Chem.* **261**, 15197-15203 (1986).
- Tanaka, K., Yoshimura, T., Ichihara, A., Kameyama, K. & Takagi, T. *J. biol. Chem.* **261**, 15204-15207 (1986).
- Zolfaghari, R., Baker, C. R. F., Canizaro, P. C., Amirgholami, A. & Behal, F. J. *Biochem. J.* **241**, 129-135 (1987).
- Hough, R., Pratt, G. & Rechsteiner, M. *J. biol. Chem.* **262**, 8303-8313 (1987).
- Kopp, F., Steiner, R., Dahlmann, B., Kuehn, L. & Reinauer, H. *Biochem. biophys. Acta* **872**, 253-260 (1986).
- Killary, A. M. & Fournier, R. E. K. *Cell* **38**, 523-534 (1984).
- Kloetzel, P.-M., Falkenburg, P.-E., Hössl, P. & Glätzer, K. H. *Exp. Cell Res.* **170**, 204-213 (1987).
- Lämmli, U. K. *Nature* **227**, 680-685 (1970).
- Kloetzel, P.-M., Johnson, R. M. & Sommerville, J. *Eur. J. Biochem.* **127**, 301-308 (1982).

Identity of the 19S 'prosome' particle with the large multifunctional protease complex of mammalian cells (the proteasome)

Andre-Patrick Arrigo*, Keiji Tanaka†
Alfred L. Goldberg‡§, & William J. Welch*

* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

† Institute for Enzyme Research, University of Tokushima Medical School, Tokushima 770, Japan

‡ Harvard Medical School, Department of Physiology and Biophysics, Boston, Massachusetts 02115, USA

There have been many reports that eukaryotic cells contain ring-shaped 19S or 20S particles which are composed of numerous polypeptide subunits ranging in size between 25 and 35 kilodaltons¹⁻¹². Because these particles seemed to copurify with inactive mRNA, they were assumed to function in regulating mRNA translation^{8,12} and hence were named 'prosome' (for 'programmed-o-some')⁸. A number of properties have been reported for these structures, including an association with specific RNA species^{3,8-12} or with certain heat-shock proteins¹⁰ and involvement in tRNA processing¹³ or aminoacyl tRNA synthesis¹. However, these proposed activities have not been supported by definitive evidence. During studies of the proteolytic systems in mammalian tissues¹⁴⁻¹⁷, we noted many similarities between these 19S particles and the high molecular weight protease complexes that are present in most or all eukaryotic cells¹⁴⁻²⁴. This (700 kilodalton) enzyme complex, designated here as LAMP for 'large alkaline multifunctional protease', contains three distinct endoproteolytic sites which function at neutral or alkaline pH and are specific for hydrolysis of proteins, hydrophobic peptides, or basic peptides¹⁴⁻¹⁹. This protease also exists in a latent form which can be activated

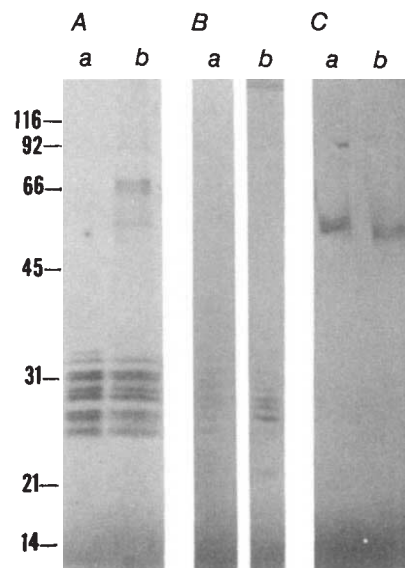


Fig. 1 Similarities in the protein composition of LAMP from human erythrocytes and the HeLa prosome. Panel A, SDS-PAGE on 15% gel: lane a, LAMP; lane b, prosome. Molecular weight markers are indicated to the left (in thousands). Panel B, the purified prosome was resolved on a 15% polyacrylamide gel, transferred to nitrocellulose and incubated with a polyclonal antibody raised against rodent LAMP (ref. 20). The primary antibody was visualized with a second horseradish peroxidase-conjugated antibody and subsequent colour development. Lane a, proteins transferred to nitrocellulose and stained with amidoblack; lane b, reaction product of prosome incubated with anti-LAMP antibody. Panel C, purified LAMP and prosome were analysed under non-denaturing conditions on a 2% acrylamide/0.5% agarose gel and stained with Coomassie blue. Lane a, LAMP; lane b, prosome. **Methods.** LAMP was purified as described²⁰. HeLa prosome was prepared from the 15,000g supernatant of hypotonically lysed HeLa cells grown in suspension by either of two methods. In the first method a 150,000g pellet from this 15,000g supernatant was purified by a series of velocity sedimentation steps through continuous sucrose gradients^{6,9}. In the second new and more rapid procedure, the prosome was purified from the 15,000g supernatant by ion-exchange and gel filtration chromatography followed by velocity sedimentation through a continuous sucrose gradient (A.A. and W.J.W., unpublished data). LAMP and prosome were solubilized in Laemmli sample buffer and the proteins separated on a 15% polyacrylamide gel and stained with Coomassie blue.

by polylysine^{14,16}, fatty acids¹⁹, or ATP^{17,25-28}. In this report, we show that the prosomes and these protease complexes are very similar or identical with respect to their size, polypeptide composition, immunological cross-reactivity, appearance in the electron microscope, radial symmetry of subunits, subcellular localization, and proteolytic activities. Therefore, the 'prosome' probably plays a critical role in intracellular protein breakdown, and we propose that it be renamed 'proteasome'.

'Prosome' were purified from HeLa cells by either of two methods which yielded particles indistinguishable in size, morphology, and protein composition (described in Fig. 1). The multifunctional protease LAMP was purified from human erythrocytes or from rat liver in its latent form, using buffers containing glycerol^{14,16}. Figure 1 (panel A) shows the polypeptide components of the purified HeLa prosome analysed by one-dimensional gel electrophoresis. A major set of polypeptides of relative molecular mass (M_r) ranging from 25,000-35,000 (25-35K) was observed in accord with previous reports¹⁻¹² (with some minor proteins of ~68-70K). Also shown are the proteins composing the LAMP from human red cells (Fig. 1a). The pattern and relative intensities of the LAMP protein bands appeared identical to those of the prosome components

§ To whom correspondence should be addressed.

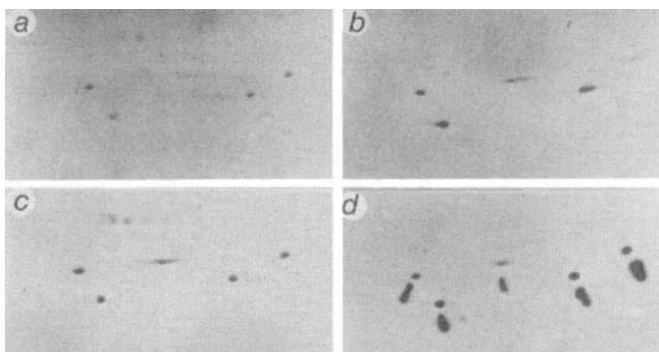


Fig. 2 Analysis of purified LAMP and prosome by 2-D gel electrophoresis. Purified unlabelled LAMP and [^3H]leucine-labelled HeLa prosome were analysed by 2-D gel electrophoresis either alone or after mixing. The acidic end of the gel is to the left. *a*, Purified LAMP proteins stained with Coomassie blue. *b*, Purified [^3H]prosomes visualized by fluorography. *c*, Mixture of purified LAMP and [^3H]prosomes stained with Coomassie blue. *d*, Fluorographs of the gel in *c* containing the mixture of LAMP and prosome. After staining with Coomassie blue, the gel was subjected to fluorography and dried. The position of the stained proteins was then indicated by radioactive ink. The film of the fluorographed gel was developed after 5 d.

(but this preparation of LAMP also contained a polypeptide of $\sim 34\text{K}$). To confirm the identity of these proteins, a polyclonal antibody raised in rabbits against the purified rat liver LAMP (ref. 14) was used in Western blot analysis of the prosome proteins after transfer of the proteins to nitrocellulose. The anti-LAMP antibody showed a positive reaction with many of the smaller proteins in the prosome (Fig. 1 panel *B*, lane *b*).

When LAMP and prosome were subjected to non-denaturing gel electrophoresis, both migrated as a single species, with little or no difference in apparent mobility (Fig. 1, panel *C*); thus, their net size and charge are very similar. The proteins comprising the prosome and LAMP were further analysed by two-dimensional gel electrophoresis using radiolabelled prosomes isolated from HeLa cells grown on [^3H]leucine. The profiles of the LAMP proteins on the 2-D gels (Fig. 2*a*) were surprisingly less complex than the 1-D results (Fig. 1*a*). Similarly, [^3H]leucine-labelled prosomes gave only 4–5 prominent spots on 2-D gels (Fig. 2*b*). When the two preparations were mixed, the radiolabelled prosomal proteins and the human erythrocyte LAMP components migrated in identical fashion (Fig. 2*c*, *d*).

Table 1 Proteolytic and peptidase activities of prosome particles from HeLa cells

Substrate	HeLa prosome (units g^{-1})	Erythrocyte multifunctional protease complex (units g^{-1})
[^{14}C]methylcasein	109	361
+ Poly-L-lysine (1 mg ml^{-1})	201	1,082
CBZ-Ala-Arg-Arg-MNA*	416	874
Glutaryl-Ala-Ala-Phe-MNA*	155	271

The purified prosome ($10 \mu\text{g ml}^{-1}$) was incubated with the substrate for 1 h as described^{14,16}. The prosome particles were isolated as in Fig. 1 legend. The protease complex purified in its latent form in glycerol is shown for comparison^{14,16}. Similar data were obtained in three independent experiments.

* MNA, methoxynaphylamine.

When the structures of the LAMP and prosome were compared in the electron microscope (Fig. 3*A* and *B*), both had a cylindrical shape, with an outer diameter of ~ 9 – 13nm surrounding a dense central zone¹⁵. (Occasionally, stacks of several rings were evident.) This 'raspberry'-like morphology has been described for the prosome from a variety of different organisms^{1–12}. To investigate the arrangement of subunits and their possible symmetry, we subjected the electron micrographs to 'Markham rotation' analysis²⁹. This technique showed that the LAMP particle had a clear 8-fold radial symmetry (Fig. 3*c*–*g*) and that the HeLa prosome had a similar, but less-pronounced, 8-fold symmetry (Fig. 3*h*–*l*). Thus there must be a precise subunit organization within the particles.

The intracellular distribution of LAMP was determined by indirect immunofluorescence analysis using an antibody raised

against LAMP (Fig. 4). In rat fibroblasts, LAMP had a punctate distribution throughout the nucleus and a somewhat more diffuse pattern within the cytoplasm. A similar distribution was observed for the prosome in the electron and light microscopes^{3,5,7,30}.

Finally, the HeLa prosome showed the enzymatic activities characteristic of the multifunctional protease (Table 1). At pH 7.6, the HeLa particles degraded [^{14}C]casein to acid-soluble products, and this process was stimulated several-fold by poly-lysine^{14,16}. The prosome also hydrolysed the basic and hydrophobic peptides, which are specific fluorogenic substrates for the other active sites on this complex^{14,16,18,19}. The relative activities differed somewhat in the two preparations, probably as a result of the different purification methods and time of storage^{16,17}. Our previous work^{14,17} has shown that three of the

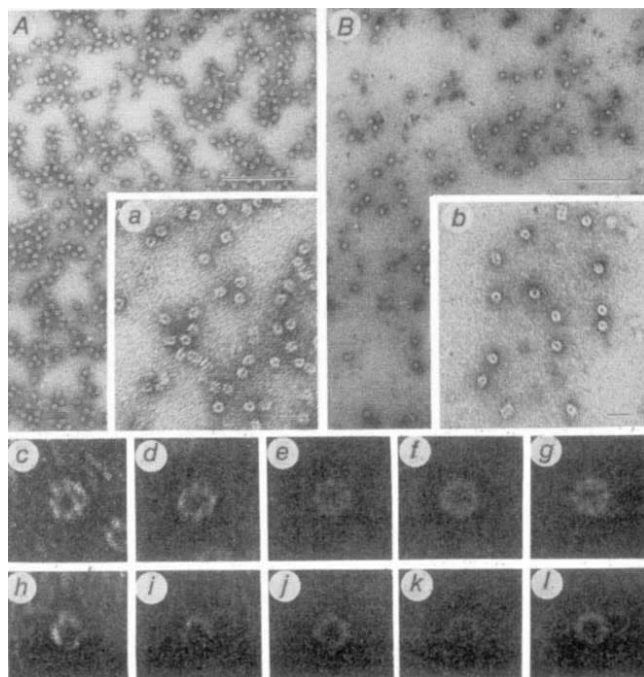


Fig. 3 Comparison of rodent LAMP and mammalian prosome by electron microscopy and Markham rotation. Panels *A*, *a* and *B*, *b*: electron micrographs after absorption of particles of LAMP (panel *A*, *a*) and HeLa prosome (panel *B*, *b*) to carbon-coated Formvar grids and negative staining with 1% aqueous uranyl acetate. Panels *A* and *B* are low-magnification survey micrographs (scale bar, 100nm) and panels *a* and *b* are enlargements (scale bar, 16 nm) of selected regions in *A* and *B*. Panels *c*–*g* and *h*–*l*: to test for radial symmetry, Markham rotation was performed by enlarging electron micrograph negatives of LAMP (panels *c*–*g*) and HeLa prosome (panels *h*–*l*) rotated $360^\circ/N$ as described²⁹. Panels *c* and *h*, N is 1; panels *d* and *i*, N is 2; panels *e* and *j*, N is 4; panels *f* and *k*, N is 7; panels *g* and *l*, N is 8.

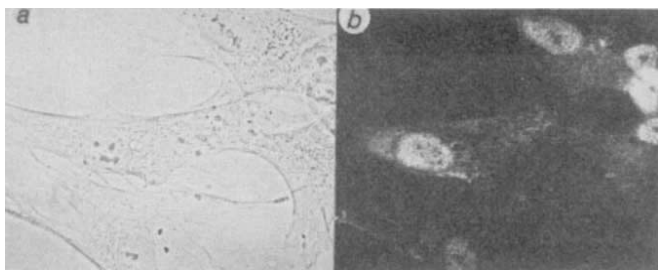


Fig. 4 Indirect immunofluorescence analysis of the intracellular distribution of LAMP in rat fibroblasts shown in phase contrast (a) and fluorescence (b). Rat embryo fibroblasts were fixed with 3.7% formalin and permeabilized by incubation with 0.2% Triton X-100. The cells were incubated with the rabbit anti-LAMP antibody, and the primary antibody visualized by subsequent incubation with a fluorescein-conjugated goat anti-rabbit antibody.

major subunits of LAMP (Fig. 1) can be labelled with [^3H]diisopropylfluorophosphate (DFP) which is evidence of their proteolytic nature. Covalent labelling of the HeLa prosome peptides with [^3H]DFP was also observed (A.A., unpublished observations).

The demonstration that the purified prosome has multiple proteolytic activity suggests that it could be involved in intracellular proteolysis. Several functions have been proposed for the prosome, including inhibition of mRNA function^{8,12}, aminoacyl-transferase activity¹, processing of transfer RNA precursors¹³, and an association with low molecular weight heat-shock proteins¹⁰. But these suggestions have generally been based on studies of impure fractions, and the ascribed activities probably resulted from contaminants. For example, we found that the 28K heat-shock protein in mammalian cells exists as a large aggregate ($\geq 15\text{S}$) which initially copurifies with the prosome³¹, but which can be separated after further purification. Although the prosome has been described as a ribonucleoprotein, our purest protease complexes from rat liver^{14,15} and the particles isolated by other groups^{6,13} do not contain RNA. Furthermore, ribonuclease treatment does not alter the enzymatic activity or morphology of the particle (unpublished observations). So it is unclear whether the presence of RNA depends on the purification procedure and if it has functional significance.

Because the terms 'prosome' and 19S RNP are uninformative and potentially misleading, we suggest these particles be renamed 'proteasomes' to indicate their enzymatic activities and probable biological function. This large enzyme was initially described²⁵⁻²⁷ as an ATP-stimulated protease that might be important in the non-lysosomal ATP-dependent pathway for protein breakdown. This conclusion is supported by its multiple peptidase activity^{14-19,23,24} and the finding that this protease in its native form is ATP dependent (J. Driscoll and A.L.G., manuscript in preparation). Like the ATP-dependent proteases found in *Escherichia coli* and mitochondria³², the proteasome functions independently of ubiquitin, and it is thus distinct from the other very large protease complex in eukaryotic cells that degrades ubiquitin-protein conjugates³³⁻³⁵. Although the two protease complexes may function together, the proteasome by itself constitutes an enzyme system that must be important in the turnover of certain cytosolic and nuclear proteins.

We thank Joe Suhan for electron microscopy and Markham rotation analysis and Dr Timothy Edmunds for the purified enzyme, also Mary Mulcahy, Phil Renna, Madeline Szadkowski for technical assistance and Aurora Scott for help in the preparation of this manuscript. This work was supported by the NIH (W.J.W. and A.L.G.), Biogen Research Corp. (A.L.G.) and a Bristol Myers Institutional Grant (A.P.A.).

Received 8 October; accepted 2 November 1987.

- Shelton, E., Kuff, E. L., Maxell, E. S. & Harrington, S. T. *J. Cell Biol.* **45**, 1-8 (1970).
- Harris, J. R. *J. molec. Biol.* **46**, 329-335 (1970).
- Narayan, S. & Round, D. E. *Nature new Biol.* **243**, 146-150 (1973).
- Smulson, M. *Expl Cell Res.* **87**, 243-258 (1974).
- Domae, N. *et al. Life Sci.* **30**, 469-477 (1982).
- Kleinschmidt, J. A., Hugle, B., Grund, C. & Franke, W. *Eur. J. Cell Biol.* **32**, 143-156 (1983).
- Hugle, B., Kleinschmidt, J. A. & Franke, W. *Eur. J. Cell Biol.* **32**, 157-163 (1983).
- Schmid, H. P. *et al. EMBO J.* **3**, 29-34 (1984).
- Arrigo, A. P., Darlix, J. L., Khandjian, E. W., Simon, M. & Spahr, P. F. *EMBO J.* **4**, 399-406 (1985).
- Schudt, C. & Kloetzel, P. M. *Dev Biol.* **110**, 65-74 (1985).
- Arrigo, A. P., Simon, M., Darlix, J. L. & Spahr, P. F. *J. molec. Biol.* **25**, 141-150 (1987).
- Martins de Sa, C. *et al. J. molec. Biol.* **187**, 479-493 (1986).
- Castano, J. G., Ormberg, R., Koster, J. G., Tobian, J. A. & Zasloff, M. *Cell* **46**, 377-387 (1986).
- Tanaka, K., Ii, K., Ichihara, A., Waxman, L. & Goldberg, A. L. *J. biol. Chem.* **261**, 15197-15203 (1986).
- Tanaka, K., Yoshimura, T., Ichihara, A., Kameyama, K. & Takagi, T. *J. biol. Chem.* **261**, 15204-15207 (1986).
- Tanaka, K., Waxman, L., Boches, A. & Goldberg, A. L. *J. biol. Chem.* (in the press).
- Tanaka, K., Waxman, L., Tawa, N. & Goldberg, A. L. *J. biol. Chem.* (in the press).
- Wilk, S. & Orlowski, M. *J. Neurochem.* **40**, 842-849 (1983).
- Dehmann, B., Kuehn, L., Rutschmann, M. & Reinauer, H. *Biochem. J.* **228**, 161-170 (1985).
- Ismail, F. & Gevers, W. *Biochim. biophys. Acta* **742**, 399-408 (1983).
- Roy, K. & Harris, H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7545-7549 (1985).
- Rivett, A. J. *J. biol. Chem.* **260**, 12600-12606 (1985).
- Edmunds, T. & Pennington, R. J. T. *Int. J. Biochem.* **14**, 701-703 (1982).
- Ishihara, S., Sano, M., Kamakura, K. & Sugita, H. *FEBS Lett.* **189**, 119-123 (1985).
- DeMartino, G. N. & Goldberg, A. L. *J. biol. Chem.* **254**, 3712-3715 (1979).
- DeMartino, G. N. *J. molec. Cell Cardiol.* **15**, 17-29 (1983).
- Boches, F. S., Klemes, Y. & Goldberg, A. L. *Fedn Proc.* **39**, 1982 (1980).
- Rose, I. A., Warms, J. V. B. & Hersko, A. J. *J. biol. Chem.* **254**, 8135-8138 (1979).
- Markham, R., Frey, S. & Hills, G. *J. Virology* **20**, 88-100 (1963).
- Kloetzel, P. M., Falkenburg, P. E., Hossli, P. & Glatzer, K. N. *Expl Cell Res.* **170**, 204-213 (1987).
- Arrigo, A. P. & Welch, W. J. *J. biol. Chem.* (in the press).
- Goldberg, A. L. *et al. Intracellular Protein Catabolism* (eds Khairallah, E. A., Bond, J. S. & Bird, J. W. C.) 603-605 (Liss, New York, 1985).
- Hough, R., Pratt, G. & Rechsteiner, M. *J. biol. Chem.* **261**, 2400-2408 (1986).
- Waxman, L., Fagan, J. M. & Goldberg, A. L. *J. biol. Chem.* **262**, 2451-2457 (1987).
- Fagan, J. M., Waxman, L. & Goldberg, A. L. *Biochem. J.* **243**, 335-343 (1987).
- Hough, R., Pratt, G. & Rechsteiner, M. *J. biol. Chem.* **262**, 8303-8313 (1987).

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse.

I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnType ☐ OCLC ILL Subsystem

☐ Other (please specify) _____

☐ I am interested in sending my order by mail.

☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name _____

Title _____

Institution/Company _____

Department _____

Address _____

City _____ State _____ Zip _____

Phone (_____) _____

Mail to: University Microfilms International
 300 North Zeeb Road, Box 91 Ann Arbor, MI 48106